

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2005, 19:13:34 ; Search time 163.5 Seconds
(without alignments)
4056.844 Million cell updates/sec

Title: US-10-021-571-4
Perfect score: 8884
Sequence: 1 MSVLISQSVINYVEENIPA.....LHAASSESTGFGBERESIL 1715

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_16Dec04:.*
1: geneseqp1980s:.*
2: geneseqp1990s:.*
3: geneseqp2000s:.*
4: geneseqp2001s:.*
5: geneseqp2002s:.*
6: geneseqp2003as:.*
7: geneseqp2003bs:.*
8: geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8884	100.0	1715	4	AAM39025 Human pol
2	8884	100.0	1715	4	AAM38993 Human pol
3	8884	100.0	1715	5	AAE25144 Human ARM
4	8884	100.0	1715	5	AAU96840 Human kid
5	8884	100.0	1715	6	AAC32128 Human cyt
6	8855	99.7	1753	6	ADA09888 Human rec
7	8830	99.4	1771	7	ADC06847 Human Kid
8	8362.5	94.1	1762	5	AAU96841 Rat Kidin
9	8362.5	94.1	1762	7	ADC06848 Rat Kidin
10	8301	93.4	1763	5	AAU80244 Rat Kidin
11	8266	93.0	1715	5	AAC25143 Rat ARMS
12	8266	93.0	1715	7	ADC06849 Rat Kidin
13	5931	66.8	1142	7	ADC06850 Kidins 22
14	5899.5	66.4	1184	7	ADC06846 Human Kid
15	3050.5	34.3	705	4	AAU75604 Human col
16	2851.5	32.1	551	4	AAE01035 Human dea
17	2791	31.4	543	4	AAU95191 Human pro
18	2638	29.7	513	4	AAE03645 Human ext
19	2503	28.2	624	4	ABG08697 Novel hum
20	2379.5	26.8	1498	4	ABB64857 Drosophil
21	2379.5	26.8	1498	5	AAE25146 Fruit fly
22	1832.5	20.6	1398	5	AAE25145 Worm ARMS
23	667	7.5	131	5	ADK36597 Novel hum
24	666	7.5	129	4	AAU94785 Human pro
25	620	7.0	120	4	AAU40811 Human pol

26	557.5	6.3	1881	7	ADD47763 Human pro
27	553	6.2	109	3	AAU56277 Human sec
28	552.5	6.2	1880	8	ADR90358 Full leng
29	547.5	6.2	986	4	ABG31785 Human ank
30	547.5	6.2	1139	7	ADC39122 Novel hum
31	547.5	6.2	1250	6	AAE33667 Human str
32	543.5	6.1	657	7	ADB64375 Human pro
33	542.5	6.1	4082	8	ABM83911 Human dia
34	539.5	6.1	1375	8	ADR09291 Human pro
35	537	6.0	1954	8	ABM83918 Human dia
36	536.5	6.0	3897	8	ABM83919 Human dia
37	536.5	6.0	3924	8	ABM83916 Human dia
38	536.5	6.0	3955	8	ABM83915 Human dia
39	536.5	6.0	3985	8	ABM83914 Human dia
40	536.5	6.0	4016	8	ABM83912 Human dia
41	536.5	6.0	4072	8	ABM83913 Human dia
42	531.5	6.0	1955	8	ABM83917 Human dia
43	530.5	6.0	1839	7	ADD27862 Human bra
44	526	5.9	1872	4	AAU79160 Human pro
45	521	5.9	4274	4	ABG00972 Novel hum

ALIGNMENTS

RESULT 1
AAM39025
ID AAM39025 standard; protein; 1715 AA.

XX AC AAM39025;
XX XX
DT 22-OCT-2001 (first entry)
DE Human polypeptide SEQ ID NO 2170.

KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.

XX Homo sapiens.

XX OS

XX PN WO200153312-A1.

XX FD 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.

XX 23-DEC-1999; 99US-00471275.

XX 21-JAN-2000; 2000US-00488725.

XX 25-APR-2000; 2000US-00552317.

XX 20-JUN-2000; 2000US-00598042.

XX 19-JUL-2000; 2000US-00620312.

XX 03-AUG-2000; 2000US-00653450.

XX 14-SEP-2000; 2000US-00662191.

XX 19-OCT-2000; 2000US-00693036.

XX 29-NOV-2000; 2000US-00727344.

(HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;

PI Zhou P, Goodrich R, Drmanac RT;

PI WPI; 2001-442253/47.

PI N-PSDB; AAI58181.

PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.

XX AAB94785 Human pro

PS Example 4; SEQ ID NO 2170; 10078pp; English.

XX	The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AA38642-AA42213) with nootropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Actin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders. Note: The sequence data for this patent did not form part of the printed specification									
XX	SQ Sequence 1715 AA;									
	Query Match	100.0%;	Score 8884;	DB 4;	Length 1715;					
	Best Local Similarity	100.0%;	Pred. No. 0;							
	Matches 1715;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;					
QY	1	MSVLISQSVINYVEENIPALKALLECKDQVDERNECGQTPLMIAAEQGNLEIVKELIKN	60							
DB	1	MSVLISQSVINYVEENIPALKALLECKDQVDERNECGQTPLMIAAEQGNLEIVKELIKN	60							
QY	61	GANCNLEDLNDWTALISASKEGHVHIVEELLKCGVNLHRDMGWTALMWACYKGRTDVV	120							
DB	61	GANCNLEDLNDWTALISASKEGHVHIVEELLKCGVNLHRDMGWTALMWACYKGRTDVV	120							
QY	121	ELLSHGANPSVTGLQYSVYPIIWAAGRHADIVHLLQNGAKVNSDKYGTTPPLVWAAR	180							
DB	121	ELLSHGANPSVTGLQYSVYPIIWAAGRHADIVHLLQNGAKVNSDKYGTTPPLVWAAR	180							
QY	181	KGHLECVKHLAMGADVQEGANSMTALIVAVKGGYTQSVKEILKRNPNVNLTKDQNTA	240							
DB	181	KGHLECVKHLAMGADVQEGANSMTALIVAVKGGYTQSVKEILKRNPNVNLTKDQNTA	240							
QY	241	LMIASKEGHTETIVODLLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRG	300							
DB	241	LMIASKEGHTETIVODLLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRG	300							
QY	301	QDNKTALYWAVEKGNATWRDILQCPDTEICTKDGTEPLIKATKRNIEVWELLDDKGA	360							
DB	301	QDNKTALYWAVEKGNATWRDILQCPDTEICTKDGTEPLIKATKRNIEVWELLDDKGA	360							
QY	361	KVSAVDKGGTPLHIAIRGRSRKLAELLRNPKDGRLLYRPNKAGETPPYINDCSHOKSIL	420							
DB	361	KVSAVDKGGTPLHIAIRGRSRKLAELLRNPKDGRLLYRPNKAGETPPYINDCSHOKSIL	420							
QY	421	TOIFGARHLSPTETDGMGLGYDLYSSALADILSEPTMQPPICVGLYAQMGSGKSFLLKKL	480							
DB	421	TOIFGARHLSPTETDGMGLGYDLYSSALADILSEPTMQPPICVGLYAQMGSGKSFLLKKL	480							
QY	481	EDEMKTFAGQIQEPLQFQSWLIIVFLTLILCCGGLGILFAFTVHNGLGIVLSFLALLYIF	540							
DB	481	EDEMKTFAGQIQEPLQFQSWLIIVFLTLILCCGGLGILFAFTVHNGLGIVLSFLALLYIF	540							
QY	541	FIVYFGGREGESWNAWVLSRLARHIGLYELLKLMFVNPPPELQTTKALPVRFPLF	600							
DB	541	FIVYFGGREGESWNAWVLSRLARHIGLYELLKLMFVNPPPELQTTKALPVRFPLF	600							
QY	601	TDYNRLSSVGGETSLAEMIALTSDACEREFGLATRLFRVFKTEDTQGGKKWKKTCCPLPS	660							
DB	601	TDYNRLSSVGGETSLAEMIALTSDACEREFGLATRLFRVFKTEDTQGGKKWKKTCCPLPS	660							
QY	661	FVIFLFIIGCIISGITLLAIFRVDPKHLTNAVLISIASVVGGLAFVNLNCRTWQVLDLSLL	720							
DB	661	FVIFLFIIGCIISGITLLAIFRVDPKHLTNAVLISIASVVGGLAFVNLNCRTWQVLDLSLL	720							
QY	721	NSQRKELHNAASKHLKSEGFMKVLKCEVELMARMAKTIDSFTQNTQTRLVVIIDGLDAC	780							

RESULT 2
AAW38993
ID AAW38993 standard; protein; 1715 AA.
XX
AC
XX AAW38993;

DT	22-OCT-2001	(first_entry)	
XX	Human polypeptide	SEQ ID NO 2138.	
XX	Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukaemia.		
XX	Homo sapiens.		
OS	WO200153312-A1.		
XX	26-JUL-2001.		
XX	26-DEC-2000; 2000WO-US034263.		
XX	23-DEC-1999; 99US-00471275.		
PR	21-JAN-2000; 2000US-00488725.		
PR	25-APR-2000; 2000US-00552317.		
PR	20-JUN-2000; 2000US-00598042.		
PR	19-JUL-2000; 2000US-00620312.		
PR	03-AUG-2000; 2000US-00653450.		
PR	14-SEP-2000; 2000US-00662191.		
PR	19-OCT-2000; 2000US-00693036.		
PR	29-NOV-2000; 2000US-00727344.		
XX	(HYSE-) HYSEQ INC.		
PA	Tang YT, Liu C, Auandi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
XX	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;		
PI	Zhou P, Goodrich R, Drmanac RT;		
PI	WPI; 2001-442253/47.		
DR	N-PSDB; AA158149.		
XX	Novel nucleic acids and polypeptides, useful for treating disorders such		
PT	as central nervous system injuries.		
XX	Example 4; SEQ ID NO 2138; 10078pp; English.		
XX	The invention relates to human nucleic acids (AA157798-AA161369) and the		
CC	encoded polypeptides (AA139642-AA142213) with nontropic,		
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful		
CC	in gene therapy. A composition containing a polypeptide or polynucleotide		
CC	of the invention may be used to treat diseases of the peripheral nervous		
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		
CC	localised neuropathies and central nervous system diseases, such as		
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the		
CC	utilisation of the activities such as: immune system suppression,		
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic		
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,		
CC	assays for receptor activity, arthritis and inflammation, leukaemia and		
CC	C.N.S disorders. Note: The sequence data for this patent did not form		
CC	part of the printed specification		
XX	Sequence 1715 AA;		
XX	Query Match	100.0%; Score 8884; DB 4; Length 1715;	
XX	Best Local Similarity	100.0%; Pred. No. 0;	
XX	Matches 1715; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	MSVLSQSVINYVEENIPALKALLEKCKVDNERNECCQTFLMTAAEQGNLEIVKELIKN	60
Db	1	MSVLSQSVINYVEENIPALKALLEKCKVDNERNECCQTFLMTAAEQGNLEIVKELIKN	60
Qy	61	GANCNLEDLDNWTALISAKSGHGHVIVBELLKCGVNLHRDGMGGWTALMWACYKGRDGV	120
Db	61	GANCNLEDLDNWTALISAKSGHGHVIVBELLKCGVNLHRDGMGGWTALMWACYKGRDGV	120

```
Db 1201 KEMNFGDWHLPFRSTVLENRNASHVVPDPRLSESSGPAHPGEPARRASHNELPHT 1260
QY 1261 ELSSTQPTTLNLFNFEELNTLGLDEGAPRHSNLNSQWQTRTPSLNSQDSSIEISKLT 1320
Db 1261 ELSSTQPTTLNLFNFEELNTLGLDEGAPRHSNLNSQWQTRTPSLNSQDSSIEISKLT 1320
QY 1321 DKVQAEYRDYRYIIQMSQLEGPGSTTISGRSSPHSTTYMGQSSGSGSIHNSLQEKG 1380
Db 1321 DKVQAEYRDYRYIIQMSQLEGPGSTTISGRSSPHSTTYMGQSSGSGSIHNSLQEKG 1380
QY 1381 KDSEPKPDDGRKGFMLKRGVIVDIYSSGVSTNDASPLDPIITEDEKSDQSGSKLLPKKXS 1440
Db 1381 KDSEPKPDDGRKGFMLKRGVIVDIYSSGVSTNDASPLDPIITEDEKSDQSGSKLLPKKXS 1440
QY 1441 SERSSLFQTDLKLKGLSGLRYQKLPSPDESGTEESDNTPLKDDKDKAEGKVERVPKSP 1500
Db 1441 SERSSLFQTDLKLKGLSGLRYQKLPSPDESGTEESDNTPLKDDKDKAEGKVERVPKSP 1500
QY 1501 EHSABIRTFIKAKELYSLDALLDKQSSDSGVRSSESSPHNSLHNEVADDSQLEKANLIE 1560
Db 1501 EHSABIRTFIKAKELYSLDALLDKQSSDSGVRSSESSPHNSLHNEVADDSQLEKANLIE 1560
QY 1561 LEDDSHSGKRGIPHSLSGLQDPIIARMSICSEDKSPSECSLIASSPEENWPACQKAYNL 1620
Db 1561 LEDDSHSGKRGIPHSLSGLQDPIIARMSICSEDKSPSECSLIASSPEENWPACQKAYNL 1620
QY 1621 NRTPTSTVTLNNSAPANRANQNFDEMGIRETSQVILRPSSSPNPTTIQENLKSMTNHR 1680
Db 1621 NRTPTSTVTLNNSAPANRANQNFDEMGIRETSQVILRPSSSPNPTTIQENLKSMTNHR 1680
QY 1681 SQSSSYRLSKDPPELHAAASSESTGFERESIL 1715
Db 1681 SQSSSYRLSKDPPELHAAASSESTGFERESIL 1715

RESULT 3
AAE25144
ID AAE25144 standard; protein; 1715 AA.
AC AAE25144;
XX
XX
DT 30-OCT-2002 (first entry)
XX
DE Human ARMS protein.
XX
KW Ankyrin repeat-rich membrane spanning protein; ARMS; neurotrophin;
KW ephrin; receptor tyrosine kinase; growth cone; neuron; neuronal cell;
KW diagnostic; imaging; human.
XX
OS Homo sapiens.
XX
FH Key
FH Region 17..390
FT /note= "Ankyrin repeat"
FT Region 354..493
FT /note= "N-terminal region"
FT Domain 496..518
FT /note= "Transmembrane domain"
FT Domain 525..546
FT /note= "Transmembrane domain"
FT Region 573..638
FT /note= "Loop between TM2 and TM3"
FT Domain 661..680
FT /note= "Transmembrane domain"
FT Domain 688..710
FT /note= "Transmembrane domain"
FT Region 749..854
FT /note= "Carboxy tail"
FT Region 940..1060
FT /note= "Carboxy tail"
FT Region 1081..1093
FT /note= "Polyproline stretch"
```

```
FT Domain 1151..1221
FT Binding-site /note= "SAM domain"
FT 1713..1715
FT /note= "PDZ binding motif"
XX
XX WO200250273-A2.
XX
XX 27-JUN-2002.
XX
XX 19-DEC-2001; 2001WO-US048603.
XX
XX 21-DEC-2000; 2000US-0256909P.
XX
XX (UYN ) UNIV NEW YORK STATE.
XX
XX Chao MV, Kong H;
XX
XX WPI; 2002-508800/54.
XX N-PSDB; AAD41036.
XX
XX Isolated ankyrin repeat-rich membrane spanning (ARMS) polypeptide that is
XX a target for phosphorylation by neurotrophin and ephrin receptor tyrosine
XX kinases, useful as a marker for growth cones.
XX
XX Claim 1; Page 103-110; 136pp; English.
XX
XX The invention relates to ankyrin repeat-rich membrane spanning (ARMS)
XX protein which is a down stream target of neurotrophin and ephrin receptor
XX tyrosine kinases. ARMS DNA is useful for visualising the growth cone of
XX neurons. ARMS protein is useful as an indicator of the biological
XX activity of neurotrophins and ephrins, as a marker for neuronal cells
XX which have the ability to undergo continued synaptic changes through
XX adult life or for the presence and distribution of ARMS in such neuronal
XX cells, or as a marker for growth cones. ARMS DNA is useful in diagnostic
XX and imaging methods. The present sequence is human ARMS protein
XX
XX Sequence 1715 AA;
XX
XX Query Match 100.0%; Score 8884; DB 5; Length 1715;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1715; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSVLISQSVINVEENIPALKLEKCKDVDERNECGOTPLMIAAEQGNLEIVKELIKN 60
Db 1 MSVLISQSVINVEENIPALKLEKCKDVDERNECGOTPLMIAAEQGNLEIVKELIKN 60
QY 61 GANCNLEDLNDNTALISASKEGHVHIVEELLKCGVNLHRDMGGWTALMWACYKGRTDVV 120
Db 61 GANCNLEDLNDNTALISASKEGHVHIVEELLKCGVNLHRDMGGWTALMWACYKGRTDVV 120
QY 121 ELLLSHGANSVTLGQYSVYPIIWAAGRGHADIVHLLLLONGAKVNSDKYGTTPPLVWAAR 180
Db 121 ELLLSHGANSVTLGQYSVYPIIWAAGRGHADIVHLLLLONGAKVNSDKYGTTPPLVWAAR 180
QY 181 KGHLECVKHLWAGADVDOEGANSMTALIVAVKGGYTQSVKELKENPNVNLTKDGNTA 240
Db 181 KGHLECVKHLWAGADVDOEGANSMTALIVAVKGGYTQSVKELKENPNVNLTKDGNTA 240
QY 241 LMIASKEGHEIVQDILLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRG 300
Db 241 LMIASKEGHEIVQDILLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRG 300
QY 301 QNKNTALYWAVKGNATWVRDILQCNPDTEICTKOGTETPLIKATKRNIEVVELLDKGA 360
Db 301 QNKNTALYWAVKGNATWVRDILQCNPDTEICTKOGTETPLIKATKRNIEVVELLDKGA 360
QY 361 KVSADVKKGDTPLHIAIRGRSRKLAELLRNPKDGLLYRPNKAGETPNYDCHSHKSL 420
Db 361 KVSADVKKGDTPLHIAIRGRSRKLAELLRNPKDGLLYRPNKAGETPNYDCHSHKSL 420
QY 421 TQIFGARHLSPTETDGMGLDYLSALADILSEPTMQPPIICVGLYAQWNGSGKSFLLKKL 480
Db 421 TQIFGARHLSPTETDGMGLDYLSALADILSEPTMQPPIICVGLYAQWNGSGKSFLLKKL 480
```

Qy	481	EDMKTFAGQOIEBLPQFSWLI	IVFLTL	LLCGIGL	LLFAFTVHP	NMGIAVS	LSF	LALI	YIP	540
Db	481	EDMKTFAGQOIEBLPQFSWLI	IVFLTL	LLCGIGL	LLFAFTVHP	NMGIAVS	LSF	LALI	YIF	540
Qy	541	FIVYFGRRREGESNNAW	LSLRLA	HIGYLE	LELLKLMFVN	PNPEL	PEOTTKAL	PVRFLP	600	
Db	541	FIVYFGRRREGESNNAW	LSLRLA	HIGYLE	LELLKLMFVN	PNPEL	PEOTTKAL	PVRFLP	600	
Qy	601	TDYNRLSSVGGETSLAEN	MIATLS	DACEREF	GLATRLFR	VRFKTED	TQGGKKK	TKCLPS	660	
Db	601	TDYNRLSSVGGETSLAEN	MIATLS	DACEREF	GLATRLFR	VRFKTED	TQGGKKK	TKCLPS	660	
Qy	661	FVIFLIIGCIISGITLLAI	FRVDPKH	LTNAVL	ISIASV	VGIAFV	LNCR	TWQV	LSLL	720
Db	661	FVIFLIIGCIISGITLLAI	FRVDPKH	LTNAVL	ISIASV	VGIAFV	LNCR	TWQV	LSLL	720
Qy	721	NSQRKRLHNAASKLHLK	SEGMVK	LKCEV	ELMARM	AKTIDSF	TQNOTRL	VVIIDGL	DAC	780
Db	721	NSQRKRLHNAASKLHLK	SEGMVK	LKCEV	ELMARM	AKTIDSF	TQNOTRL	VVIIDGL	DAC	780
Qy	781	EODKVLQMLDTRVRL	PSKGPPIA	PASDPH	IIKAIN	QNLS	VLRD	SNINGH	DYMR	840
Db	781	EODKVLQMLDTRVRL	PSKGPPIA	PASDPH	IIKAIN	QNLS	VLRD	SNINGH	DYMR	840
Qy	841	LPVFLNSRGLSNARKFL	VTSATNG	DVPCSD	TGTQED	ADRRV	SONS	LGEM	TKLSK	900
Db	841	LPVFLNSRGLSNARKFL	VTSATNG	DVPCSD	TGTQED	ADRRV	SONS	LGEM	TKLSK	900
Qy	901	RRDYRRRRQMTITROM	SFDLTK	LVTE	DWFS	DISP	QTMRL	LLNI	SV	960
Db	901	RRDYRRRRQMTITROM	SFDLTK	LVTE	DWFS	DISP	QTMRL	LLNI	SV	960
Qy	961	FNWDRLASWINLTQW	PVRTSW	LILYLE	ETEGIP	DQMT	LTKI	YERIS	KN	1020
Db	961	FNWDRLASWINLTQW	PVRTSW	LILYLE	ETEGIP	DQMT	LTKI	YERIS	KN	1020
Qy	1021	EIDGDIRNFVFLSR	PTVLVAR	DKVFL	PCTV	NLDPK	LREIIAD	VRARE	QIS	1080
Db	1021	EIDGDIRNFVFLSR	PTVLVAR	DKVFL	PCTV	NLDPK	LREIIAD	VRARE	QIS	1080
Qy	1081	PPLFLHEGPPRAPSGY	SQPPSC	VSCTSF	NGP	FAGGV	SPHSS	YVS	GMT	1140
Db	1081	PPLFLHEGPPRAPSGY	SQPPSC	VSCTSF	NGP	FAGGV	SPHSS	YVS	GMT	1140
Qy	1141	SGPAPGPVLLNS	LVND	AVCE	KLKQI	IEGL	DOSML	POYCT	TIK	1200
Db	1141	SGPAPGPVLLNS	LVND	AVCE	KLKQI	IEGL	DOSML	POYCT	TIK	1200
Qy	1201	KEMNNFGDWHLFR	STVLE	MNASH	VVPED	PRFL	SESS	GPAP	HGPAR	1260
Db	1201	KEMNNFGDWHLFR	STVLE	MNASH	VVPED	PRFL	SESS	GPAP	HGPAR	1260
Qy	1261	ELSSQTPYTLN	FLNFS	FEEL	NLTGL	DEGA	PHSN	LSWQS	QTR	1320
Db	1261	ELSSQTPYTLN	FLNFS	FEEL	NLTGL	DEGA	PHSN	LSWQS	QTR	1320
Qy	1321	DKVQAEYRD	AYRE	IAQMS	QLEG	PGSTTIS	GRSP	SHST	YMG	1380
Db	1321	DKVQAEYRD	AYRE	IAQMS	QLEG	PGSTTIS	GRSP	SHST	YMG	1380
Qy	1381	KDSPKPPDDGK	KSFLMK	RGD	VIVYSS	SGVSN	DAS	PLDPT	TEE	1440
Db	1381	KDSPKPPDDGK	KSFLMK	RGD	VIVYSS	SGVSN	DAS	PLDPT	TEE	1440
Qy	1441	SERSSLFQTD	LKXG	SLRYQ	KLP	SDDE	SGTES	SDNT	PL	1500
Db	1441	SERSSLFQTD	LKXG	SLRYQ	KLP	SDDE	SGTES	SDNT	PL	1500
Qy	1501	EHSAPERTFT	KAKEY	LS	DALLDK	OSD	SGVR	SESS	PNH	1560
Db	1501	EHSAPERTFT	KAKEY	LS	DALLDK	OSD	SGVR	SESS	PNH	1560

Qy	1561	LEDDSHSGKRGIPHSLSGLODPIIARMWICSDKKSFPSECSLIASSPEENWPACQKAYNL	1620
Db	1561	LEDDSHSGKRGIPHSLSGLODPIIARMWICSDKKSFPSECSLIASSPEENWPACQKAYNL	1620
Qy	1621	NRTPSTVTLLNNAPANRANONFDEMGIRETSQVILRPSSSSNPPTTIQENLUKSMTHKR	1680
Db	1621	NRTPSTVTLLNNAPANRANONFDEMGIRETSQVILRPSSSSNPPTTIQENLUKSMTHKR	1680
Qy	1681	SQSSSYRLSKDPPPELHAAASSESTGFGEERESIL	1715
Db	1681	SQSSSYRLSKDPPPELHAAASSESTGFGEERESIL	1715
RESULT 4			
AAU96840			
XX	ID	AAU96840 standard; protein; 1715 AA.	
AC	AAU96840;		
XX	30-JUL-2002 (first entry)		
DT	Human kidneys220 protein.		
DE			
XX	Kidins220; kinase D interacting substrate of 22kDa; cytostatic;		
KW	neuroprotective; gene therapy; protein kinase D; PKD; cancer;		
KW	neurodegenerative disease; glioblastoma multiforme; prostate cancer;		
KW	human.		
XX			
OS	Homo sapiens.		
XX			
XX	W0200220786-A2.		
FN			
XX	14-MAR-2002.		
PD			
XX			
XX	06-SEP-2001; 2001WO-GB003977.		
PF			
XX			
XX	06-SEP-2000; 2000US-0230449P.		
PR			
XX	(IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.		
PA			
XX	Schiavo G, Iglesias T;		
PI			
XX	WPI; 2002-371879/40.		
DR	N-PSDB; ABKS1221.		
DR			
XX			
XX			
PT	Polypeptide kinase D interacting substrate of 220 kDa and polynucleotide,		
PT	useful for identifying modulators useful in treating cancer and		
PT	neurodegenerative diseases.		
XX			
XX	Claim 33; Fig 11; 228pp; English.		
XX			
CC	The invention relates to polypeptide comprising a kinase D interacting		
CC	substrate of 220 kDa (Kidins220) from rat and human or their fragment,		
CC	variant or fusion provided that the protein is not the polypeptide		
CC	encoded by GenBank Accession No. AB033076. Also included are a		
CC	polynucleotide encoding Kidins220 provided that polynucleotide is not any		
CC	one of the clones corresponding to the 61 GenBank Accession Nos. given in		
CC	the specification; an expression/replicable vector comprising the		
CC	polynucleotide; a recombinant host cell containing the polynucleotide or		
CC	vector; an anti-Kidins220 antibody (used in the preparation of Kidins220		
CC	; an agent (A1) which modulates activity of Kidins220, protein kinase D		
CC	(PKD) or interaction between PKD and Kidins220; a mutant animal		
CC	transgenic for kidins220; the use of an agent capable of detecting the		
CC	expression of kidins220 gene in manufacture of a diagnostic reagent for		
CC	diagnosing or prognosing cancer or for monitoring the progression of		
CC	cancer in a patient; a compound comprising a moiety that selectively		
CC	binds to Kidins220 or its variant and another moiety; and a nucleic acid		
CC	molecule encoding the compound. The antibody is useful for modulating the		
CC	activity of Kidins220, where the antibody prevents a region of kidins220		
CC	interacting from another protein, or reduces the ability of Kidins220 to		
CC	bind to adenosine triphosphate (ATP). The kidins220 polynucleotide is		
CC	useful for identifying an agent which modulates the promoter activity of		
CC	the polynucleotide, and Kidins220 is useful for identifying an agent		

CC which modulates activity of Kidins220, protein kinase D (PKD) and the
 CC interaction between PKD with Kidins220. Kidins220 or its polynucleotide
 CC is useful in medicine, e.g. using gene therapy or for diagnosing or
 CC monitoring progression of cancer in a patient. The agent is useful in
 CC manufacture of medicament, for use in treatment of neurodegenerative
 CC disease, and the agent which inhibits function of Kidins220 gene or its
 CC product is useful for treating cancer which is glioblastoma multiforme or
 CC prostate cancer. The compound is useful for imaging cancer in an
 CC individual, and for diagnosing or prognosing, and also for treating an
 CC individual. The present sequence represents Human Kidins220
 XX
 SQ Sequence 1715 AA;

Query Match 100.0%; Score 8884; DB 5; Length 1715;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1715; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSVLISQSVINYVEENIPALKALLECKDVKDERNECGQTPLMAAEOGNLEIVKELIKN 60
 DB 1 MSVLISQSVINYVEENIPALKALLECKDVKDERNECGQTPLMAAEOGNLEIVKELIKN 60

QY 61 GANCNLEDLNTALISASKEGHVHIVEELLKCGVNLHRDMGGMWTALMWACYKGRDGV 120
 DB 61 GANCNLEDLNTALISASKEGHVHIVEELLKCGVNLHRDMGGMWTALMWACYKGRDGV 120

QY 121 ELLLSHGPNPVTGLQSVYPIIWAAGRGHADIHVLLONGAKVNSDKYGTTPVWAAR 180
 DB 121 ELLLSHGPNPVTGLQSVYPIIWAAGRGHADIHVLLONGAKVNSDKYGTTPVWAAR 180

QY 181 KGHLECVKHLAMGADVQGGANSMALIVAVKGGYTQSVKEILKRNPNVNLTKDGNTA 240
 DB 181 KGHLECVKHLAMGADVQGGANSMALIVAVKGGYTQSVKEILKRNPNVNLTKDGNTA 240

QY 241 LMIASKEGHTIEIVQDLIDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDRG 300
 DB 241 LMIASKEGHTIEIVQDLIDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDRG 300

QY 301 QDNKTALYWAVERGNATWVRDILQCPDTEICTKQGETPLIKATKRNIEVVELLDKGA 360
 DB 301 QDNKTALYWAVERGNATWVRDILQCPDTEICTKQGETPLIKATKRNIEVVELLDKGA 360

QY 361 KVSADVKKGTPLHAIIRGRSKLAELLRNPNKDGRLLYRPNKAGTTPYNDICSHQSKIL 420
 DB 361 KVSADVKKGTPLHAIIRGRSKLAELLRNPNKDGRLLYRPNKAGTTPYNDICSHQSKIL 420

QY 421 TQIFGARHLSPTETDGMGLGYDLYSSALADILSEPTMQPPICVGLYAQWGSGKSFLLKKL 480
 DB 421 TQIFGARHLSPTETDGMGLGYDLYSSALADILSEPTMQPPICVGLYAQWGSGKSFLLKKL 480

QY 481 EDEMKTFAGQOIEPLFQFSLWLIIVFLTLCCGGGLLFAFTVHNLGIAVLSFLALLYIF 540
 DB 481 EDEMKTFAGQOIEPLFQFSLWLIIVFLTLCCGGGLLFAFTVHNLGIAVLSFLALLYIF 540

QY 541 FIVIFGGRREGSGSNWVLSRLARHIGYLELLKJLMPVNPPELPEQTTKALPVRELF 600
 DB 541 FIVIFGGRREGSGSNWVLSRLARHIGYLELLKJLMPVNPPELPEQTTKALPVRELF 600

QY 601 TDYNRLSSVSGGETSLAEMIATLSDACEREFGLATRLFRVFKTEDTQGGKKWKTKCCLPS 660
 DB 601 TDYNRLSSVSGGETSLAEMIATLSDACEREFGLATRLFRVFKTEDTQGGKKWKTKCCLPS 660

QY 661 FVIFLFIIGCIIISGITLLAI FVRDPKHLTVNAVLISIASVVGIAFVNLNCRTHWQVLDL 720
 DB 661 FVIFLFIIGCIIISGITLLAI FVRDPKHLTVNAVLISIASVVGIAFVNLNCRTHWQVLDL 720

QY 721 NSORKRLHNAASKHLKSKSGFMKVLKCEVELMARMAKTIIDSTFQNTQRLVLIIDGLDAC 780
 DB 721 NSORKRLHNAASKHLKSKSGFMKVLKCEVELMARMAKTIIDSTFQNTQRLVLIIDGLDAC 780

QY 781 EQDKVLQMLDTPVRLFSKGFPIAFASDPHIIKAINQNLSVLRDSNINGHDYMNIVH 840
 DB 781 EQDKVLQMLDTPVRLFSKGFPIAFASDPHIIKAINQNLSVLRDSNINGHDYMNIVH 840

QY 841 LPVFLNSRGLSNARKFLVTSATNGDVPVCSDDTTGGIOEDADRRVSONSLGEMTKLGSKTALN 900
 DB 841 LPVFLNSRGLSNARKFLVTSATNGDVPVCSDDTTGGIOEDADRRVSONSLGEMTKLGSKTALN 900

QY 901 RRDYTRRRQMQRTITRQMSFDLTLLVTEWDFSDISQPMRRLNIVSVTGRLLRANQIS 960
 DB 901 RRDYTRRRQMQRTITRQMSFDLTLLVTEWDFSDISQPMRRLNIVSVTGRLLRANQIS 960

QY 961 FNDWRLASWINLQEQWPFYRTSWLLYLETEGIPDMTLKTIYERISKNIPTTKDVEPLL 1020
 DB 961 FNDWRLASWINLQEQWPFYRTSWLLYLETEGIPDMTLKTIYERISKNIPTTKDVEPLL 1020

QY 1021 EIDGDIRNEFEVLSRTPVLVAROVKFLPCTVNLDPKLEIADVRAAREQISIGGLAY 1080
 DB 1021 EIDGDIRNEFEVLSRTPVLVAROVKFLPCTVNLDPKLEIADVRAAREQISIGGLAY 1080

QY 1081 PPLPLHEGPPRAPSGYSQPPSVCSSTSFNGPAGGVVSPQPHSSYSGMTGQHPFYNRG 1140
 DB 1081 PPLPLHEGPPRAPSGYSQPPSVCSSTSFNGPAGGVVSPQPHSSYSGMTGQHPFYNRG 1140

QY 1141 SGAPGPGVLLNSLVNDVAVCEKLKQIEGLDQSMLEPOYCTTIKKANINGRVLACNIDELK 1200
 DB 1141 SGAPGPGVLLNSLVNDVAVCEKLKQIEGLDQSMLEPOYCTTIKKANINGRVLACNIDELK 1200

QY 1201 KEMNNFQDWHLFRSTVLEMRNAESHVVPEDPRFLSESSSGPAPGEPARRASHNELPHT 1260
 DB 1201 KEMNNFQDWHLFRSTVLEMRNAESHVVPEDPRFLSESSSGPAPGEPARRASHNELPHT 1260

QY 1261 ELSSTQPTPLNFSPEELNTLGLDEGAPRHSNLSWSQSTRTPSLSLNSQDSSIEISKLT 1320
 DB 1261 ELSSTQPTPLNFSPEELNTLGLDEGAPRHSNLSWSQSTRTPSLSLNSQDSSIEISKLT 1320

QY 1321 DKVQAEYRDAYREYIAQMSQLEGGPGSTTISGRSPHSTYTMGQSSSGSIHNSLQEK 1380
 DB 1321 DKVQAEYRDAYREYIAQMSQLEGGPGSTTISGRSPHSTYTMGQSSSGSIHNSLQEK 1380

QY 1381 KQSEKPDGGRKFLMKRGDVIDYSSGVSTNDASPLDPIITEDEKSDGSKLLPGKKS 1440
 DB 1381 KQSEKPDGGRKFLMKRGDVIDYSSGVSTNDASPLDPIITEDEKSDGSKLLPGKKS 1440

QY 1441 SERSSIFQTDLKLKSGSLRYQKLPSEDESGTEESDNTPLLKDDKDKAEGKVERVPKSP 1500
 DB 1441 SERSSIFQTDLKLKSGSLRYQKLPSEDESGTEESDNTPLLKDDKDKAEGKVERVPKSP 1500

QY 1501 EHSABPRTFIKAYELSDALLDKDSDSGVRSSESSPNHSLHNEVADDQLEKANLIE 1560
 DB 1501 EHSABPRTFIKAYELSDALLDKDSDSGVRSSESSPNHSLHNEVADDQLEKANLIE 1560

QY 1561 LEDDSSHGKRGIPHSLSGLQDPIIARMSICSEDKSKSPSECSLIASSPEENWPACQAYNL 1620
 DB 1561 LEDDSSHGKRGIPHSLSGLQDPIIARMSICSEDKSKSPSECSLIASSPEENWPACQAYNL 1620

QY 1621 NRTPTSTVTLNNSAPANRANQNFDEMEGIRETSQVILRPSSSPNPTTIQENLKSMTKRR 1680
 DB 1621 NRTPTSTVTLNNSAPANRANQNFDEMEGIRETSQVILRPSSSPNPTTIQENLKSMTKRR 1680

QY 1681 SQRSSYTRLSKDPPELHAAASSESTGFGERESIL 1715
 DB 1681 SQRSSYTRLSKDPPELHAAASSESTGFGERESIL 1715

RESULT 5
 AAE32128
 ID AAE32128 standard; protein; 1715 AA.
 XX
 AC AAE32128;
 XX
 DT 24-MAR-2003 (first entry)
 XX
 DE Human cytoskeleton-associated protein, CSAP-26.
 DE Human; cytoskeleton-associated protein; CSAP-26; atherosclerosis; cancer;
 KW gene therapy.
 KW

Db 421 TOIFGARHLSPTETDGMGLDYLSALADILSEPTMQPPICVGLYAQWGSKSFLLK 480
Qy 481 EDEMTKFAQOIEPLFQFQSWLIVFTLLCGGLGALLFAFVHPNGLIAVLSFLALLYIF 540
Db 481 EDEMTKFAQOIEPLFQFQSWLIVFTLLCGGLGALLFAFVHPNGLIAVLSFLALLYIF 540
Qy 541 FIVIVFGRRREGESNNWAWLSTRLARHIGYELLLKLMFVNPPPEQTTKALPVRFLF 600
Db 541 FIVIVFGRRREGESNNWAWLSTRLARHIGYELLLKLMFVNPPPEQTTKALPVRFLF 600
Qy 601 TDYNRLSSVGGTSLAEMTIATLSDACEREFGLATRLFRVFKTEQCKKWKTCCLPS 660
Db 601 TDYNRLSSVGGTSLAEMTIATLSDACEREFGLATRLFRVFKTEQCKKWKTCCLPS 660
Qy 661 FVIFLFIIGCIISGITLLAIFRVDPKHLTVNVLISIASVVGGLAFVLCNRTWQVLDL 720
Db 661 FVIFLFIIGCIISGITLLAIFRVDPKHLTVNVLISIASVVGGLAFVLCNRTWQVLDL 720
Qy 721 NSQKRLHNAASKLHLKSEGFPMKVLKCEVELMARMAKTIDSFQTNQTRLVVIDGLDAC 780
Db 721 NSQKRLHNAASKLHLKSEGFPMKVLKCEVELMARMAKTIDSFQTNQTRLVVIDGLDAC 780
Qy 781 EODKVLQMLDTRVRLFSKGPPIAFASDPHIIIIKAINQNLSVLSDNSINGHDYMRNI 840
Db 781 EODKVLQMLDTRVRLFSKGPPIAFASDPHIIIIKAINQNLSVLSDNSINGHDYMRNI 840
Qy 841 LPVFLNSRGLSNARKFLVTSATNGDVPCSDTTGIOEDADRRVSONSLGEMTKLGSKTALN 900
Db 841 LPVFLNSRGLSNARKFLVTSATNGDVPCSDTTGIOEDADRRVSONSLGEMTKLGSKTALN 900
Qy 901 RRDYRRRQMQRTITRQMSFDLTLLVTEDEWFSDISPQTMRLNLNIVSVTGRLLRANQIS 960
Db 901 RRDYRRRQMQRTITRQMSFDLTLLVTEDEWFSDISPQTMRLNLNIVSVTGRLLRANQIS 960
Qy 961 FNWDLASWINLTEQWPRTSLIILYBETEGIPQWTLKTIYERISQNIPTTKDVEPLL 1020
Db 961 FNWDLASWINLTEQWPRTSLIILYBETEGIPQWTLKTIYERISQNIPTTKDVEPLL 1020
Qy 1021 EIDGDIRNFEVFLSRTFVLVARDVKVFLPCTVNDPKLREIADVRAAREQISIGGLAY 1080
Db 1021 EIDGDIRNFEVFLSRTFVLVARDVKVFLPCTVNDPKLREIADVRAAREQISIGGLAY 1080
Qy 1081 PPLPLHEGPAPSGYSPVSCSSTFNGPPAGVSPQPHSSYSYSGMTGPQHPFYNR - 1139
Db 1081 PPLPLHEGPAPSGYSPVSCSSTFNGPPAGVSPQPHSSYSYSGMTGPQHPFYNR 1140
Qy 1140 -----GSGPAPGPVVLNLSLNDVAVCEK 1162
Db 1141 FPAPLYLYTRYPPGSGOHLISRPVSKTSLPRDQNNGLSGSPAGPVPVLLNSLNDVAVCEK 1200
Qy 1163 LKQIEGLDQSMPLPOYCTTIKKANINGRVLAOCNIDELKEMNFGDWHLFRSTVLEWRN 1222
Db 1201 LKQIEGLDQSMPLPOYCTTIKKANINGRVLAOCNIDELKEMNFGDWHLFRSTVLEWRN 1260
Qy 1223 AESHVVPEDPRLSSESSGAPHGPARASHNELPHTLSQTPYTLNFSFEELNTLGL 1282
Db 1261 AESHVVPEDPRLSSESSGAPHGPARASHNELPHTLSQTPYTLNFSFEELNTLGL 1320
Qy 1283 DEGAPRHNLSWQSOTRTPSLSSLSNSQDSISIEISKLTKDQVQAYRDYRVIYIAQMSOLE 1342
Db 1321 DEGAPRHNLSWQSOTRTPSLSSLSNSQDSISIEISKLTKDQVQAYRDYRVIYIAQMSOLE 1380
Qy 1343 GPGGTTISGRSSPHSTHYMGSSGGSIHNLQEKGDSEPKDCKRKFPLMKRGDVI 1402
Db 1381 GPGGTTISGRSSPHSTHYMGSSGGSIHNLQEKGDSEPKDCKRKFPLMKRGDVI 1440
Qy 1403 DYSSSGVSTNDASPLDPIITEDEKSDQSGSKLLPGKSSERSLFTQDLKLGSLRYQK 1462
Db 1441 DYSSSGVSTNDASPLDPIITEDEKSDQSGSKLLPGKSSERSLFTQDLKLGSLRYQK 1500
Qy 1463 LPSDEDESGTBSNDTPLLKDDKORKAEGKVERVPKSPHSAEPIRTPIKAKYLSDAL 1522
Db 1501 LPSDEDESGTBSNDTPLLKDDKORKAEGKVERVPKSPHSAEPIRTPIKAKYLSDAL 1560

Qy 1523 DKKSSDSGVSRSSSPNHLHNEVADDSQLEKANLIELEDDSHSGKRGIPHSLSGLQDP 1582
Db 1561 DKKSSDSGVSRSSSPNHLHNEVADDSQLEKANLIELEDDSHSGKRGIPHSLSGLQDP 1620
Qy 1583 IIA RMSICSEDKKSPSECSLIASSPEENWPAQKAYNLNRTPTSTVTLNNSAPANRANQN 1642
Db 1621 IIA RMSICSEDKKSPSECSLIASSPEENWPAQKAYNLNRTPTSTVTLNNSAPANRANQN 1680
Qy 1643 FDEMEGIRETSQVILRPSSSPNPTTIQENILKSMTHKESQSSYTRLSKDPPELHAAASS 1702
Db 1681 FDEMEGIRETSQVILRPSSSPNPTTIQENILKSMTHKESQSSYTRLSKDPPELHAAASS 1740
Qy 1703 ESTGFGERESIL 1715
Db 1741 ESTGFGERESIL 1753

RESULT 7
ADC06847

ID ADC06847 standard; protein; 1771 AA.

XX ADC06847;

XX AC ADC06847;

XX DT 18-DEC-2003 (first entry)

XX XX Human Kidins220Pc protein AB033076.

XX DE cytostatic; prostate cancer; breast; gene therapy; transgenic; human;

XX KW Kidins220Pc; kinase D-interacting substrate of 220kDa; chromosome 2p25.1.

XX OS Homo sapiens.

XX XX WO2003064599-A2.

XX PD 07-AUG-2003.

XX PF 24-JAN-2003; 2003WO-US001943.

XX PR 25-JAN-2002; 2002US-00054935.

XX PR 14-FEB-2002; 2002US-0356130P.

XX PR 22-MAR-2002; 2002US-00102946.

XX PR 08-APR-2002; 2002US-00117229.

XX PR 14-MAY-2002; 2002US-00144198.

XX PR 19-JUL-2002; 2002US-00197824.

XX XX (ORIG-) ORIGENE TECHNOLOGIES INC.

XX PI Sun Z, Li X, Jay G, Kovacs KF, Fan W, Shu Y;

XX XX WPI; 2003-679495/64.

XX PT New isolated polynucleotide related to cancer genes, useful for

XX PT detecting, diagnosing, staging, monitoring, prognosticating, preventing

XX PS or treating cancers, e.g. breast and prostate cancers.

XX PS Disclosure; Fig 20; 128pp; English.

XX CC The invention relates to a novel isolated polynucleotide comprising a
XX CC differentially-regulated mammalian cancer gene. The polynucleotides of
XX CC the invention demonstrate cytostatic activity and are differentially
XX CC expressed in prostate cancer. The polynucleotide, polypeptides and
XX CC methods of the invention may be useful for detecting, diagnosing,
XX CC staging, monitoring, prognosticating, preventing or treating cancers,
XX CC particularly breast and prostate cancers. Furthermore, the invention may
XX CC be utilised during gene therapy procedures or in the production of
XX CC transgenic animals. The current sequence is that of the prostate cancer-
XX CC related protein of the invention. The current sequence is not fully
XX CC legible within the specification; the complete sequence was obtained from
XX CC GenBank.

SQ Sequence 1771 AA;

Query Match		99.4%;	Score 8830;	DB 7;	Length 1771;	
Best Local Similarity		96.7%;	Pred. No. 0;			
Matches 1714;		Conservative	0;	Mismatches	0;	Indels 58; Gaps 2;
QY	1	MSVLISQSVINYVEEENIPALKALLECKVDKVDNERNECGQTPLMTAAEQGNLEIVKELIKN	60			
DB	1	MSVLISQSVINYVEEENIPALKALLECKVDKVDNERNECGQTPLMTAAEQGNLEIVKELIKN	60			
QY	61	GANCNLEDLQNTALISASKEGHVHIVVEELLKCGVNLNHRDMGGWTALMWACVKGRTDVV	120			
DB	61	GANCNLEDLQNTALISASKEGHVHIVVEELLKCGVNLNHRDMGGWTALMWACVKGRTDVV	120			
QY	121	ELLISHGANPVTGLQYVVPITWAAGRHADIVHLLLQNGAKVNCSDKYGTTPPLVWAAR	180			
DB	121	ELLISHGANPVTGL-YSVTPITWAAGRHADIVHLLLQNGAKVNCSDKYGTTPPLVWAAR	179			
QY	181	KGHLECVKHLAMGADVDOEGANSMTALIIVAVKGGYTQSVKEILKRNPNVNLTKDQNTA	240			
DB	180	KGHLECVKHLAMGADVDOEGANSMTALIIVAVKGGYTQSVKEILKRNPNVNLTKDQNTA	239			
QY	241	LMIASKEGHTIEIVODLLDAGTYNIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRG	300			
DB	240	LMIASKEGHTIEIVODLLDAGTYNIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRG	299			
QY	301	QDNKTALYWAKEGNATWVRDILQCNPDTEICTKQGETPLIKATQWNIIEVVELLDKGA	360			
DB	300	QDNKTALYWAKEGNATWVRDILQCNPDTEICTKQGETPLIKATQWNIIEVVELLDKGA	359			
QY	361	KVSAVDKGGTPTLHIAIRGSRKLAELLRNPKDGRLLYRPNKAGETPPYINDCSHQKSI	420			
DB	360	KVSAVDKGGTPTLHIAIRGSRKLAELLRNPKDGRLLYRPNKAGETPPYINDCSHQKSI	419			
QY	421	TOIFGARHLSPETDGMGLYDLYSSALADILSEPTWQPPICVGLYAQWSGSGKSFLLK	480			
DB	420	TOIFGARHLSPETDGMGLYDLYSSALADILSEPTWQPPICVGLYAQWSGSGKSFLLK	479			
QY	481	EDEMKTFAGQOIREPLFOFSLWLVFLTLGCGGLLPAFTVHNLGIAVSLFALLIYIF	540			
DB	480	EDEMKTFAGQOIREPLFOFSLWLVFLTLGCGGLLPAFTVHNLGIAVSLFALLIYIF	539			
QY	541	FIVYFGRREGESWNAWVLSRLARHIGYLELLKLMFVNPPPELPEQTTKALPVRFLF	600			
DB	540	FIVYFGRREGESWNAWVLSRLARHIGYLELLKLMFVNPPPELPEQTTKALPVRFLF	599			
QY	601	TDYNRLSSVGGETSLAEMIATLSDACERBFGFLATRLFRVFKTEDTGKKKWKTCCLPS	660			
DB	600	TDYNRLSSVGGETSLAEMIATLSDACERBFGFLATRLFRVFKTEDTGKKKWKTCCLPS	659			
QY	661	FVIFLFIIGCTIIGITLALIPRVDPKHLTVNAVLIISTASVVGGLAVLNCRTWQVLDL	720			
DB	660	FVIFLFIIGCTIIGITLALIPRVDPKHLTVNAVLIISTASVVGGLAVLNCRTWQVLDL	719			
QY	721	NSQRKRLHNAASKLHLKLBEGFMKVLKCEVELMARMAKTTIDSFQNTQRLVLIIDGLDAC	780			
DB	720	NSQRKRLHNAASKLHLKLBEGFMKVLKCEVELMARMAKTTIDSFQNTQRLVLIIDGLDAC	779			
QY	781	EQDKVLMQDTRVLPFKSGPPIAIPASDPHIIKAINQNLSVLRDSNINCHDYMRNIVH	840			
DB	780	EQDKVLMQDTRVLPFKSGPPIAIPASDPHIIKAINQNLSVLRDSNINCHDYMRNIVH	839			
QY	841	LPVFLNSRGLSNARKFLVTSATNGDVPVCSDTTGTQEDADRRVQNSIGEMTKLGSKTALN	900			
DB	840	LPVFLNSRGLSNARKFLVTSATNGDVPVCSDTTGTQEDADRRVQNSIGEMTKLGSKTALN	899			
QY	901	RRDTYRRQQRITITROMSFDLTKLLVTEDFWSDISQPMRRLINIVSVTGRLLRANQIS	960			
DB	900	RRDTYRRQQRITITROMSFDLTKLLVTEDFWSDISQPMRRLINIVSVTGRLLRANQIS	959			
QY	961	FNWDLASWINLTQWPYRTSWLLIYLEETEGIPDQWTLKTIYVERISKNIPTTKDVEPLL	1020			
DB	960	FNWDLASWINLTQWPYRTSWLLIYLEETEGIPDQWTLKTIYVERISKNIPTTKDVEPLL	1019			
QY	1021	EIDGDIRNFEVLSRTPVLVARDVKVFLPCTVNLDPKREIITADVRAAREQISIGGLAY	1080			
DB	1020	EIDGDIRNFEVLSRTPVLVARDVKVFLPCTVNLDPKREIITADVRAAREQISIGGLAY	1079			
QY	1081	PLPLHEGPPRAPSGYQPPSCVSTSPNGPPAGGVVSPQPHSSYYSGMTGPQHPFYN--	1138			
DB	1080	PLPLHEGPPRAPSGYQPPSCVSTSPNGPPAGGVVSPQPHSSYYSGMTGPQHPFYNRP	1139			
QY	1139	-----RSGSP	1143			
DB	1140	FPAPYLYTRYYPGGSHLISRPVKTSLPRDQNGLEVIKEDAAEGLSPPTDSSRSGSP	1199			
QY	1144	APGPVLLNSLNDVACEKLKQIEGLDQSMLEPOYCTTIKKANINGRVLQAQCNIDELKEM	1203			
DB	1200	APGPVLLNSLNDVACEKLKQIEGLDQSMLEPOYCTTIKKANINGRVLQAQCNIDELKEM	1259			
QY	1204	NMFGDWHLFRSTVLEMRNAESHVVPEDPRFLSESSSGPAPGEPARRASHNELPHTELS	1263			
DB	1260	NMFGDWHLFRSTVLEMRNAESHVVPEDPRFLSESSSGPAPGEPARRASHNELPHTELS	1319			
QY	1264	SQTPYTLNFSFEELNTLGLDEGNPHSNLSWOSQTRTPSLSSLSNSQDSSIEISKLTDKV	1323			
DB	1320	SQTPYTLNFSFEELNTLGLDEGNPHSNLSWOSQTRTPSLSSLSNSQDSSIEISKLTDKV	1379			
QY	1324	QAEYRDAYREYIAQMSQLEGGPGSTTISGRSSPHSYTYMGQSSGSIHSNLQEKKGDS	1383			
DB	1380	QAEYRDAYREYIAQMSQLEGGPGSTTISGRSSPHSYTYMGQSSGSIHSNLQEKKGDS	1439			
QY	1384	EPKPDGRKSFMLMKRGDVIDYSSGVSTNDASPLDPIITEDEKSDQSGSKLLPGKKSSER	1443			
DB	1440	EPKPDGRKSFMLMKRGDVIDYSSGVSTNDASPLDPIITEDEKSDQSGSKLLPGKKSSER	1499			
QY	1444	SSLFOTDLKKGSLRYOKLPSDEDESGTEESDNTPLKDDKDKAEGKVERVPKSPHS	1503			
DB	1500	SSLFOTDLKKGSLRYOKLPSDEDESGTEESDNTPLKDDKDKAEGKVERVPKSPHS	1559			
QY	1504	AEPIRTFIKAKYLSLDALDKDSDSGVRSSESSPNHSLHNEVADDSQLEKANLIELED	1563			
DB	1560	AEPIRTFIKAKYLSLDALDKDSDSGVRSSESSPNHSLHNEVADDSQLEKANLIELED	1619			
QY	1564	DSHSGKRGIPHSLSGLQDPIIARMSICSEBKKSPSCSLIASPENNWPACQKAYNLNRT	1623			
DB	1620	DSHSGKRGIPHSLSGLQDPIIARMSICSEBKKSPSCSLIASPENNWPACQKAYNLNRT	1679			
QY	1624	PSVTTLNNSAPANRANQNFDEMEGIRETSQVTLRSPSSPNPTTIONENLKSMTKRQR	1683			
DB	1680	PSVTTLNNSAPANRANQNFDEMEGIRETSQVTLRSPSSPNPTTIONENLKSMTKRQR	1739			
QY	1684	SSYTRLSKDPPELHAAASSESTGFGEERESIL	1715			
DB	1740	SSYTRLSKDPPELHAAASSESTGFGEERESIL	1771			
RESULT 8						
AAU96841						
ID	AAU96841	standard; protein; 1762 AA.				
XX	AAU96841;					
XX	09-SEP-2004 (revised)					
DT	30-JUL-2002 (first entry)					
XX	Rat kidins220 protein from Genbank AAC35195.					
DE	Kidins220; kinase D interacting substrate of 22kDa; cytosstatic;					
XX	neuroprotective; gene therapy; protein kinase D; PKD; cancer; rat;					
KW	neurodegenerative disease; glioblastoma multiforme; prostate cancer.					
XX	Rattus sp.					
OS	Unidentified.					
XX	WO200220786-A2.					
PN	14-MAR-2002.					
XX						
PD						

XX 06-SEP-2001; 2001WO-GB003977.
PF
XX 06-SEP-2000; 2000US-0230449P.
PR
XX (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
PA
XX Schiavo G, Iglesias T;
PI
XX WPI; 2002-371879/40.
DR
XX Polypeptide kinase D interacting substrate of 220 kDa and polynucleotide,
PT useful for identifying modulators useful in treating cancer and
PT neurodegenerative diseases.
XX
PS Disclosure; Fig 18; 228pp; English.
XX
CC The invention relates to polypeptide comprising a kinase D interacting
CC substrate of 220 kDa (Kidins220) from rat and human or their fragment,
CC variant or fusion provided that the protein is not the polypeptide
CC encoded by GenBank Accession No. AB033076. Also included are a
CC polynucleotide encoding Kidins220 provided that polynucleotide is not any
CC one of the clones corresponding to the 61 GenBank Accession Nos. given in
CC the specification; an expression/replicable vector comprising the
CC polynucleotide; a recombinant host cell containing the polynucleotide or
CC vector; an anti-Kidins220 antibody (used in the preparation of Kidins220)
CC ; an agent (A1) which modulates activity of Kidins220, protein kinase D
CC (PKD) or interaction between PKD and Kidins220; a mutant animal
CC transgenic for Kidins220; the use of an agent capable of detecting the
CC expression of Kidins220 gene in manufacture of a diagnostic reagent for
CC diagnosing or prognosing cancer or for monitoring the progression of
CC cancer in a patient; a compound comprising a moiety that selectively
CC binds to Kidins220 or its variant and another moiety; and a nucleic acid
CC molecule encoding the compound. The antibody is useful for modulating the
CC activity of Kidins220, where the antibody prevents a region of Kidins220
CC interacting from another protein, or reduces the ability of Kidins220 to
CC bind to adenosine triphosphate (ATP). The Kidins220 polynucleotide is
CC useful for identifying an agent which modulates the promoter activity of
CC the polynucleotide, and Kidins220 is useful for identifying an agent
CC which modulates activity of Kidins220, protein kinase D (PKD) and the
CC interaction between PKD with Kidins220. Kidins220 or its polynucleotide
CC is useful in medicine, e.g. using gene therapy or for diagnosing or
CC monitoring progression of cancer in a patient. The agent is useful in
CC manufacture of medicament, for use in treatment of neurodegenerative
CC disease, and the agent which inhibits function of Kidins220 gene or its
CC product is useful for treating cancer which is glioblastoma multiforme or
CC prostate cancer. The compound is useful for imaging cancer in an
CC individual, and for diagnosing or prognosing, and also for treating an
CC individual. The present sequence represents Rat Kidins220 protein from
CC Genbank accession number AAG35185
CC
CC Revised record issued on 09-SEP-2004 : Correction to Organism field
XX
SQ Sequence 1762 AA;

Query Match 94.1%; Score 8362.5; DB 5; Length 1762;
Best Local Similarity 91.6%; Pred. No. 0;
Matches 1616; Conservative 47; Mismatches 50; Indels 51; Gaps 4;

Qy 1 MSVLISQSVINYVEENIPALKALLECKVDNERECQTPLMIAAQQNLEIVKELIKN 60
Db 1 MSVLISQSVINYVEENIPALKALLECKVDNERECQTPLMIAAQQNLEIVKELIKN 60

Qy 61 GANCNLEDLDNWTALISAKGHHVIEELLKCGVNLNHRDMGCGTALMWACYGRDYY 120
Db 61 GANCNLEDLDNWTALISAKGHHVIEELLKCGVNLNHRDMGCGTALMWACYGRDYY 120

Qy 121 ELLLSHGANGPSVTGLQVSYVPIIWAAGRGHADI VHLHLLQNGAKVNCSDKYGTTPLVWAAR 180
Db 121 ELLLSHGANGPSVTGLQVSYVPIIWAAGRGHADI VHLHLLQNGAKVNCSDKYGTTPLVWAAR 179

Qy 181 KGHLECVKHLAMGADVQEGANSMTALIVAVKGYTQSVKEILKRNPNVNLTKDGNNTA 240
Db 181 KGHLECVKHLAMGADVQEGANSMTALIVAVKGYTQSVKEILKRNPNVNLTKDGNNTA 240

Db 180 KGHLECVKHLAMGADVQEGANSMTALIVAVKGYTQSVKEILKRNPNVNLTKDGNNTA 239
Qy 241 LMIASKEGHTIEIVODLLDAGTYVNI PDRSGDTVLIGAVRGHVEIVRALLQYADIDIRG 300
Db 240 LMIASKEGHTIEIVODLLDAGTYVNI PDRSGDTVLIGAVRGHVEIVRALLQYADIDIRG 299
Qy 301 QDNKTALYWAVEKGNATWVRDILQCNPDTEICTKDGTEPLIKATKMRNIEVVELLDKGA 360
Db 300 QDNKTALYWAVEKGNATWVRDILQCNPDTEICTKDGTEPLIKATKMRNIEVVELLDKGA 359
Qy 361 KVSADVKKGTPLHIAIRGRSRKLAELLLRNPKDGRLLYRNKAGETPYNTDCSHQSKIL 420
Db 360 KVSADVKKGTPLHIAIRGRSRKLAELLLRNPKDGRLLYRNKAGETPYNTDCSHQSKIL 419
Qy 421 TQIFGARHLSPTETDGMVGYDLYSSALADILSPTMQPPICVGLYQAQWGGKGFLLKKL 480
Db 420 TQIFGARHLSPTETDGMVGYDLYSSALADILSPTMQPPICVGLYQAQWGGKGFLLKKL 479
Qy 481 EDEMKTFAGQIEPLPQFQSWLIVELTLLCGGLGLFAFTVHPNLGIAVSLFLALYIF 540
Db 480 EDEMKTFAGQIEPLPQFQSWLIVELTLLCGGLGLFAFTVHPNLGIAVSLFLALYIF 539
Qy 541 FIVYFGRREGESNNWAWLSTRLARHIGYLELLKLMFVNPPELPEQTTKALPVRELF 600
Db 540 FIVYFGRREGESNNWAWLSTRLARHIGYLELLKLMFVNPPELPEQTTKALPVRELF 599
Qy 601 TDYNRLSSVGETSLAEMIATLSDACEREGFLATRLFRVFKTEDTGKKKWKTCCLPS 660
Db 600 TDYNRLSSVGETSLAEMIATLSDACEREGFLATRLFRVFKTEDTGKKKWKTCCLPS 659
Qy 661 FVIFLFTIGCTITLLAIFRVDPKHLTVNAVLISIASVVGFLAVNLNCRTWQVLDL 720
Db 660 FVIFLFTIGCTITLLAIFRVDPKHLTVNAVLISIASVVGFLAVNLNCRTWQVLDL 719
Qy 721 NSQRKLHNAASKLHKLSGFGFMKLVCEVELMARMARKTIDSTFQNTQRLVVIIDGLDAC 780
Db 720 NSQRKLHNAASKLHKLSGFGFMKLVCEVELMARMARKTIDSTFQNTQRLVVIIDGLDAC 779
Qy 781 EQDKVQLMDTVRVLFSGGPFIAPFASDPHIIKAINQNLNLSVLRDNINGHDMRNIVH 840
Db 780 EQDKVQLMDTVRVLFSGGPFIAPFASDPHIIKAINQNLNLSVLRDNINGHDMRNIVH 839
Qy 841 LPVFLNSRGLSNARKFLVTSATNGDVPFSDTTGQEDADRRVSQNSLGEMTKLASKTALN 900
Db 840 LPVFLNSRGLSNARKFLVTSATNGDITCSDTTGQEDTDRVSQNSLGEMTKLASKTALN 899
Qy 901 RRDYRRRQMQRTITROMSFDLTKLLVTEWDFSDISQPMRRLNIVSVTRLLRANOIS 960
Db 900 RRDYRRRQMQRTITROMSFDLTKLLVTEWDFSDISQPMRRLNIVSVTRLLRANOIT 959
Qy 961 FNDRLASWNLTEQWPYRTSWLLLYLEETEGIPDMQTLTIYERISKNIPTTKDVEPLL 1020
Db 960 FNDRLASWNLTEQWPYRTSWLLLYLEETEGIPDMQTLTIYERISKNIPTTKDVEPLL 1019
Qy 1021 EIDGDIRNFVFLSSRTPVLVARDVKVFLPCTVNLNPKLREIIADVRAAREQISIGGLAY 1080
Db 1020 EIDGDIRNFVFLSSRTPVLVARDVKVFLPCTVNLNPKLREIIADVRAAREQINIGGLAY 1079
Qy 1081 PPLPLHGGPPRAGSYQPPSVCSSTFNGFPAGGVSPQPHSSYYSGMTQPHFPYNR - 1139
Db 1080 PPLPLHGGPPRAGSYQPPSVCSSTFNGFPAGGVSPQPHSSYYSGMTQPHFPYNR 1139
Qy 1140 -----GSG-----PAPGPVLL 1151
Db 1140 FPAPLYTPRYPGGSOHLISRSVKTSLPRDQNGLPDGSFGFNKQQAQVATGSSLL 1199
Qy 1152 NSLNVDAVCBKJQIEGLDQSMPLQYCTTIKKANINGRVLQAQCNIDELKKEMNNFQDWH 1211
Db 1200 SSMVTDVVCEKLRQIEGLDQSMPLQYCTTIKKANINGRVLQAQCNIDELKKEMANNFGDWH 1259
Qy 1212 LFRSTVLEMRNASHVVPEDPRFLSESSSGPAPGEPARRASHNELPHTLSSQTPYTLN 1271
Db 1260 LFRSMVLEMRSVESQVVPEDPRFLNENSSAPVPHGESARRSSHTLPTLSSQTPYTLN 1319

QY 1272 FSEELNTLGLDGPRLHNSQWQOTRTPSPSLSSNSQDSSIEISKLTDKVQAEYRDAY 1331
DB 1320 FSEELNTLGLDGPRLHNSQWQOTRTPSPSLSSNSQDSSIEISKLTDKVQAEYRDAY 1379
QY 1332 REYIAQMSQLEGGPGSTTISGRSPHSTYYMGQSSSGGSIHNSLEQEKGDSEPKPDQGR 1391
DB 1380 REYIAQMSQLEGGPGSTTISGRSPHSTYYMGQSSSGGSIHNSLEQEKGDSEPKPDQGR 1439
QY 1392 KSFMLKRGDVIDYSSGVSNDASPLDPIERDEKSDQSGSKLLPGKKSERSLSLFTDL 1451
DB 1440 KSFMLKRGDVIDYSSGVSNDASPLDPIERDEKSDQSGSKLLPGKKSERSLSLFTDL 1499
QY 1452 KLGSGRLRYOKLPSDEDESSTESDNTPLLKDKDKRAEKGKVERVPKSPHSAEPIRTFI 1511
DB 1500 KLGSGRLRYOKLPSDEDESSTESDNTPLLKDKDKRAEKGKVERVPKSPHSAEPIRTFI 1559
QY 1512 KAKLEYLSALLDKKSDSGVRSSESPHNSHLENEVADDSQLEKANLIELEDDSHSGKRG 1571
DB 1560 KAKLEYLSALLDKKSDSGVRSSESPHNSHLENEVADDSQLEKANLIELEDDSHSGKRG 1619
QY 1572 IPHSLGLOPPIIARMSICSEDKSPSECSLIASSPEENWPACQKAYNLNRTPTVTLNN 1631
DB 1620 MPHSLGLOPPIIARMSICSEDKSPSECSLIASSPEESWPACQKAYNLNRTPTVTLNN 1679
QY 1632 NSAPANRANQNFEMEGIRETSOVLIRPSSPNPTTIONENLKSMTKRSQORSYTRLSK 1691
DB 1680 NTAPTNRANQNFEMEGIRETSOVLIRPSSPNPTTIONENLKSMTKRSQORSYTRLSK 1739
QY 1692 DPPELHAASSESTGFEERESIL 1715
DB 1740 DASELH-AAASSESTGFEERESIL 1762

RESULT 9

ADCO6848
ID ADC06848 standard; protein; 1762 AA.
AC ADC06848;
XX
XX
DT 18-DEC-2003 (first entry)
XX
DE Rat Kidins220Pc protein AF2339045.
XX
XX
KW cytostatic; prostate cancer; breast; gene therapy; transgenic; rat;
KW Kidins220Pc; kinase D-interacting substrate of 220KDa.
XX
OS Rattus sp.
XX
PN W02003064599-A2.
XX
PD 07-AUG-2003.
XX
XX 24-JAN-2003; 2003WO-US001943.
XX
XX 25-JAN-2002; 2002US-00054935.
PR 14-FEB-2002; 2002US-0356130P.
PR 22-MAR-2002; 2002US-00102946.
PR 08-APR-2002; 2002US-00117229.
PR 14-MAY-2002; 2002US-00144198.
PR 19-JUL-2002; 2002US-00197824.
XX
XX (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
XX Sun Z, Li X, Jay G, Kovacs KF, Fan W, Shu Y;
XX WPI; 2003-679495/64.
XX
XX New isolated polynucleotide related to cancer genes, useful for
PT detecting, diagnosing, staging, monitoring, prognosticating, preventing
PT or treating cancers, e.g. breast and prostate cancers.
XX
XX Disclosure; Fig 20; 128pp; English.

XX The invention relates to a novel isolated polynucleotide comprising a
CC differentially-regulated mammalian cancer gene. The polynucleotides of
CC the invention demonstrate cytostatic activity and are differentially
CC expressed in prostate cancer. The polynucleotide, polypeptides and
CC methods of the invention may be useful for detecting, diagnosing,
CC staging, monitoring, prognosticating, preventing or treating cancers,
CC particularly breast and prostate cancers. Furthermore, the invention may
CC be utilised during gene therapy procedures or in the production of
CC transgenic animals. The current sequence is that of the prostate cancer-
CC related protein of the invention. The current sequence is not fully
CC legible within the specification; the complete sequence was obtained from
CC GenBank.

XX Sequence 1762 AA;

QY 1 MSVLSQSIVNVEENIPALKKLEKCKVDNERECGQTPLMIAAEQGNLEIVKELIKN 60
DB 1 MSVLSQSIVNVEENIPALKKLEKCKVDNERECGQTPLMIAAEQGNLEIVKELIKN 60
QY 61 GANCNLELDNNNTALISASKEGHVHIVEELLKCGVNLEHRDMGWTALMWACYKGRDVV 120
DB 61 GANCNLELDNNNTALISASKEGHVHIVEELLKCGVNLEHRDMGWTALMWACYKGRDVV 120
QY 121 ELLLSHGANSPTGQYQYVYPIIWAAGRGHADIHLLQNGAKVNSDKYGTTPLYWAAR 180
DB 121 ELLLSHGANSPTGQYQYVYPIIWAAGRGHADIHLLQNGAKVNSDKYGTTPLYWAAR 179
QY 181 KGHLECVKHLAMGADVDOEGANSMTALIVAKGGYTQSVKEILKNPNVNLTKDGNTA 240
DB 180 KGHLECVKHLAMGADVDOEGANSMTALIVAKGGYTQSVKEILKNPNVNLTKDGNTA 239
QY 241 LMTASKEGHEIIVQDLDLAGTVYNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDRG 300
DB 240 LMTASKEGHEIIVQDLDLAGTVYNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDRG 299
QY 301 QDNKTALYWAKEGNAWVRDILQCNPDTEICTKDGTEPLIKATKMRNIEVVELLDKGA 360
DB 300 QDNKTALYWAKEGNAWVRDILQCNPDTEICTKDGTEPLIKATKMRNIEVVELLDKGA 359
QY 361 KYSADVKKGDTPLHAIIRGSRKLAELLRNPKDGBLLYRPNKAGETPNIDCSHOKSIL 420
DB 360 KYSADVKKGDTPLHAIIRGSRKLAELLRNPKDGBLLYRPNKAGETPNIDCSHOKSIL 419
QY 421 TQIFGARHLSPTETDGMGLGYDLYSSALADILSEPTMQPPICVGLYAQSGSGKSFLLKKL 480
DB 420 TQIFGARHLSPTETDGMGLGYDLYSSALADILSEPTMQPPICVGLYAQSGSGKSFLLKKL 479
QY 481 EDEMKTFAQQQIEPLFQFQSWLIVFLTLCCGGIGLGLFAFTVHPNLGIAVSLSLALLYIF 540
DB 480 EDEMKTFAQQQIEPLFQFQSWLIVFLTLCCGGIGLGLFAFTVHPNLGIAVSLSLALLYIF 539
QY 541 FIVYFGRREGESMNWAWLSTRLARHIGYLELLKLMFVNPPELPEQTTKALPVRFLF 600
DB 540 FIVYFGRREGESMNWAWLSTRLARHIGYLELLKLMFVNPPELPEQTTKALPVRFLF 599
QY 601 TDYNRLSSVGGETSLSAEMIATLSDACEREFGLATRLFRVFKTETDQGGKKWKKTCCPLS 660
DB 600 TDYNRLSSVGGETSLSAEMIATLSDACEREFGLATRLFRVFKTETDQGGKKWKKTCCPLS 659
QY 661 FVIFLFIIGCIISGITLLAIFRVDPKHLTVNAVLISIASVGLAFVLCNCRNTMWQVLDL 720
DB 660 FVIFLFIIGCIISGITLLAIFRVDPKHLTVNAVLISIASVGLAFVLCNCRNTMWQVLDL 719
QY 721 NSORKELHNAASKHLKLGSEGMKVUKCEVELMARMAKTIIDSTFQNTQRLVLIIDGLDAC 780
DB 720 NSORKELHNAASKHLKLGSEGMKVUKCEVELMARMAKTIIDSTFQNTQRLVLIIDGLDAC 779
QY 781 EQDKVLQMLDTRVRLFSKGFPIAIPASDPHIIKAINQNLNSVLRDSNINGHDMYRNIVH 840

Query Match 94.1%; Score 8362.5; DB 7; Length 1762;

Best Local Similarity 91.6%; Pred. No. 0;

Matches 1616; Conservative 47; Mismatches 50; Indels 51; Gaps 4;

Db 780 EODKVLQMLDTRVLFSPGPIAI PASDPHIIKAINQNLNSVL RSDNSINGHDYMRNIVH 839
Qy 841 LPVFLNSGLSNARKFLVTSATNGDVPCSDTTGIGQEDADRRVSONSLGEMTKLGSKTALN 900
Db 840 LPVFLNSGLSNARKFLVTSATNGDITCSDTTGIGQEDDRRVSONSLGEMTKLGSKTALN 899
Qy 901 RRDYRRRQMQRTITROMSFDLT KLLVTEDFWSDISPTQMRRLNIVSVTGRLLRANQIS 960
Db 900 RRDYRRRQMQRTITROMSFDLT KLLVTEDFWSDISPTQMRRLNIVSVTGRLLRANQIT 959
Qy 961 FNDRLASWINLTQWPFRTSWLILYLEETEGIPQWMTLKIYERISKNIPTTKDVEPLL 1020
Db 960 FNDRLASWINLTQWPFRTSWLILYLEETEGIPQWMTLKIYERISKNIPTTKDVEPLL 1019
Qy 1021 EIDGDIRNFVFLSRTPVLVARDVKVFLPCTVNLDPKRLRIIADVRAAREQISIGGLAY 1080
Db 1020 EIDGDIRNFVFLSRTPVLVARDVKVFLPCTVNLDPKRLRIIADVRAAREQINIGGLAY 1079
Qy 1081 PPLPLHEGPPRAPGYSQPPSVCSSTSFNGPFPAGGVSPQPHSSYSGMTGQHPFYNR- 1139
Db 1080 PPLPLHEGPPRAPGYSQPPSVCSSTSFNGPFPAGGVSPQPHSSYSGMTGQHPFYNRP 1139
Qy 1140 -----GSG-----PAPGPVVLL 1151
Db 1140 FPAPLYTPRYPGSQHLISRSSVKTSLPRDONNGLPCDSGFNKRQQAAPVATGSSLLL 1199
Qy 1152 NSLNVDAVCEKLOI EGDQSMLOPYCTTIKKANINGRVLQACNIDELKEMNNFGDWH 1211
Db 1200 SSMVTVDVCEKRLQIEGUDQSMLOPYCTTIKKANINGRVLQACNIDELKEMNNFGDWH 1259
Qy 1212 LFRSTVLEMRNAESHVPEPRFLSESSSGPAPGEPARRASHNELPHELSSQTPYTLN 1271
Db 1260 LFRSWVLEMRVSVQVPEPRFLNENSSAPVPHGESARRSHTLPELTELSSQTPYTLN 1319
Qy 1272 PSFELNLTGLDEGAPRHNSLWSQOTRTTSLSLNSQDSSEIISKLTDKVQAEYRDAY 1331
Db 1320 PSFELNLTGLDEGAPRHNSLWSQOTRTTSLSLNSQDSSEIISKLTDKVQAEYRDAY 1379
Qy 1332 REYIAQMSQLEGGPGSTTISGRSSPHSTYVMGQSSSGSIHSLNLEQEKDSEPKPDGR 1391
Db 1380 REYIAQMSQLEGGPGSTTISGRSSPHSTYVMGQSSSGSIHSLNLEQEKDSEPKPDGR 1439
Qy 1392 KSFLMKRGDVIDYSSSGVSTNDASPLDPIITEDEKSDQSGSKLLPGKKSSRSLSLFTDL 1451
Db 1440 KSFLMKRGDVIDYSSSGVSTNDASPLDPIITEDEKSDQSGSKLLPGKKSSRSLSLFTDL 1499
Qy 1452 KLKSGGLRYQKLPDEDESGTEESDNTPLLDKDKORKAAGKVERVPKSPHSAPIRTFI 1511
Db 1500 KLKSGGLRYQKLPDEDESGTEESDNTPLLDKDKORKAAGKVERVPKSPHSAPIRTFI 1559
Qy 1512 KAKEVLSDALLDKDDSDSGVRSSESSPNHSLHNEVADDSOLEKANLIELEDHSGKRG 1571
Db 1560 KAKEVLSDALLDKDDSDSGVRSSESSPNHSLHNEVADDSOLEKANLIELEDHSGKRG 1619
Qy 1572 IPHSLGLQDPIIARMSICSDKSPSCSLIASSPEENWPACQKAYNLNRTPTVTILNN 1631
Db 1620 MPHSLGLQDPIIARMSICSDKSPSCSLIASSPEENWPACQKAYNLNRTPTVTILNN 1679
Qy 1632 NSAPANRANQNFDEWEGIRETSQVILRPSSSPNPTTTONENLKMTHKRSQSSYTRLK 1691
Db 1680 NTAPTNRANQNFDEWEGIRETSQVILRPSSSPNPTTTONENLKMTHKRSQSSYTRLK 1739
Qy 1692 DPPELHAAASSESTGFGEERESIL 1715
Db 1740 DASELH-AASSESTGFGEERESIL 1762

RESULT 10
AAU80244
ID AAU80244 standard; protein; 1763 AA.
XX
AC AAU80244;

XX 09-SEP-2004 (revised)
DT 30-JUL-2002 (first entry)
XX Rat kidins220 protein.
XX Kidins220; kinase D interacting substrate of 22kDa; cytostatic;
KW neuroprotective; gene therapy; protein kinase D; PKD; cancer; rat;
KW neurodegenerative disease; glioblastoma multiforme; prostate cancer.
XX Rattus sp.
OS Unidentified.
XX WO200220786-A2.
XX 14-MAR-2002.
XX 06-SEP-2001; 2001WO-GB003977.
XX 06-SEP-2000; 2000US-0230449P.
XX (IMCR) IMPERIAL CANCER RES TECHNOLOGY LTD.
XX Schiavo G, Iglesias T;
XX WPI; 2002-371879/40.
XX N-PSDB; ABK51214.
XX Polypeptide kinase D interacting substrate of 220 kDa and polynucleotide,
PT useful for identifying modulators useful in treating cancer and
XX neurodegenerative diseases.
XX Claim 33; Fig 1; 228pp; English.
XX The invention relates to polypeptide comprising a kinase D interacting
CC substrate of 220 kDa (Kidins220) from rat and human or their fragment,
CC variant or fusion provided that the protein is not the polypeptide
CC encoded by GenBank Accession No. AB033076. Also included are a
CC polynucleotide encoding Kidins220 provided that polynucleotide is not any
CC one of the clones corresponding to the 61 GenBank Accession Nos. given in
CC the specification; an expression/replicable vector comprising the
CC polynucleotide; a recombinant host cell containing the polynucleotide or
CC vector; an anti-Kidins220 antibody (used in the preparation of Kidins220)
CC ; an agent (A1) which modulates activity of Kidins220, protein kinase D
CC (PKD) or interaction between PKD and Kidins220; a mutant animal
CC transgenic for Kidins220; the use of an agent capable of detecting the
CC expression of Kidins220 gene in manufacture of a diagnostic reagent for
CC diagnosing or prognosing cancer or for monitoring the progression of
CC cancer in a patient; a compound comprising a moiety that selectively
CC binds to Kidins220 or its variant and another moiety; and a nucleic acid
CC molecule encoding the compound. The antibody is useful for modulating the
CC activity of Kidins220, where the antibody prevents a region of Kidins220
CC interacting from another protein, or reduces the ability of Kidins220 to
CC bind to adenosine triphosphate (ATP). The Kidins220 polynucleotide is
CC useful for identifying an agent which modulates the promoter activity of
CC the polynucleotide, and Kidins220 is useful for identifying an agent
CC which modulates activity of Kidins220, protein kinase D (PKD) and the
CC interaction between PKD with Kidins220. Kidins220 or its polynucleotide
CC is useful in medicine, e.g. using gene therapy or for diagnosing or
CC monitoring progression of cancer in a patient. The agent is useful in
CC manufacture of medicament, for use in treatment of neurodegenerative
CC disease, and the agent which inhibits function of Kidins220 gene or its
CC product is useful for treating cancer which is glioblastoma multiforme or
CC prostate cancer. The compound is useful for imaging cancer in an
CC individual, and for diagnosing or prognosing, and also for treating an
CC individual. The present sequence represents Rat Kidins220
XX
XX Revised record issued on 09-SEP-2004 : Correction to Organism field
XX Sequence 1763 AA;
XX Query Match 93.4%; Score 8301; DB 5; Length 1763;
XX Best Local Similarity 90.9%; Pred. No. 0;

Matches 1603; Conservative 47; Mismatches 64; Indels 50; Gaps 2;			
QY	1	MSVLISQSVINYVEENIPALKALLEKCKVDVDRNECGQTPLMIAAEQGNLEIVKELIKN	60
Db	1	MSVLISQSVINYVEENIPALKALLEKCKVDVDRNECGQTPLMIAAEQGNVEIVKELLKN	60
QY	61	GANCNLEDLNTWALISASKEGHVHVIEELLKGVNLEHRDMGWTALMWACVKGRTDVV	120
Db	61	GANCNLEDLNTWALISASKEGHIHVIEELLKSGALEHRDMGWTALMWACVKGRTDVV	120
QY	121	ELLISHGANPSVTGLOYSVYPIIWAAGRHADIVHLLLQNGAKVNCSDKYGTTPLYWAAR	180
Db	121	ELLISHGANPSVTGLOYSVYPIIWAAGRHADIVHLLLQNGAKVNCSDKYGTTPLYWAAR	180
QY	181	KHLECVKHLANGADVQEGANSMTALIVAVKGYTQSVKEILKNPNVNLTKDKGNTA	240
Db	181	KHLECVKHLANGADVQEGANSMTALIVAVKGYTQSVKEILKNPNVNLTKDKGNTA	240
QY	241	LMTASKEGHEIVQDLDLDAGTYNIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRG	300
Db	241	LMTASKEGHEIVQDLDLDAGTYNIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRG	300
QY	301	QDNKTALYWAVEKGNATWVDILQCNPDTEICTKOGETPLIKATKMRNIEVVELLDKGA	360
Db	301	QDNKTALYWAVEKGNATWVDILQCNPDTEICTKOGETPLIKATKMRNIEVVELLDKGA	360
QY	361	KVSAVDKKGTPHIAIRGSRKLAELLRNPKDGRLLYRPNKAGETPNYIDCSHOKSIL	420
Db	361	KVSAVDKKGTPHVAIRGSRRLAELLRNPKDGRLLYRPNKAGETPNYIDCSHOKSIL	420
QY	421	TOIFGARHSPPTDGMGLGYDYSSALADILSEPTMOPPICVGLYAQMGSGSFLKKL	480
Db	421	TOIFGARHSPPTDGMGLGYDYSSALADILSEPTMOPPICVGLYAQMGSGSFLKKL	480
QY	481	EDEMKTFAGQIQEPLFOFQSWLIVFLTLCCGGLGLFAFTVHPNGLIAVSLSLALLYIF	540
Db	481	EDEMKTFAGQIQEPLFOFQSWLIVFLTLCCGGLGLFAFTVDTNLAIASLSFLALYIF	540
QY	541	FIVIFYGRRREGSWMWVLSRLARHIGYLBELLKMFVNPPPEOTTKALPVRFLLF	600
Db	541	FIVIFYGRRREGSWMWVLSRLARHIGYLBELLKMFVNPPPELAEQTTKALPVRFLLF	600
QY	601	TDYRNLSSVCGETSLEAMITLSDACEREGFLATRLFRVFKETDQKKKKWKTCLPS	660
Db	601	TDYRNLSSVCGETSLEAMITLSDACEREGFLATRLFRVLRTEESQGGKKWKTCLPS	660
QY	661	FVIFLFTIGCIIIGITLLAIFRVDPKHLTVNAVLISIASVVGFLAVLNCRTWQVLDLIL	720
Db	661	FVIFLFTIGCIIAGITLLAIFRVDPKHLTVNAVLISIASVVGFLAVLNCRTWQVLDLIL	720
QY	721	NSQKRLHNAASKHLKSGEGMKVLKCEVELMARMAKTIDSFTQNTQLRVLIIDGLDAC	780
Db	721	NSQKRLHSAASKHLKSGEGMKVLKCEVELMARMAKTIDSFTQNTQLRVLIIDGLDAC	780
QY	781	EQDKVLQMLDVTVLFSKGFPIAFASDPHIIKAINQNLSVLRDSNNGHDMYNIHVH	840
Db	781	EQDKVLQMLDVTVLFSKGFPIAFASDPHIIKAINQNLSVLRDSNNGHDMYNIHVH	840
QY	841	LPVFLNSRGLSNARKFLVTSATNGDVPSCDTTGIQEDADRRVSQNSIGEMTKLGSKTALN	900
Db	841	LPVFLNSRGLSNARKFLVTSATNGDITCSDTTGTQEDTDRRVQNSIGEMTKLGSKTALN	900
QY	901	RRTYRRRQMQRTITROMSPDLTKLVTEWDFSDISQPNWRRLLNIVSVTGRLLRANQIS	960
Db	901	RRTYRRRQMQRTITROMSPDLTKLVTEWDFSDISQPTMRRLLNIVSVTGRLLRANQIT	960
QY	961	FNNDRLASWNLTEQPYRTSWLIVLEETEGIPDQMTLKTIVIERISKNPPTTKDVEPLL	1020
Db	961	FNNDRLASWNLTEQPYRTSWLIVLEETEGIPDQMTLKTIVIERISKNPPTTKDVEPLL	1020
QY	1021	EIDGDIRNFVFLSSRTPVLVARDVKVFLPCTVNLDPKLEIADVRAAREQISIGGLAY	1080
Db	1021	EIDGDIRNFVFLSSRTPVLVARDVKTFLLPCTVNLDPKLEIADVRAAREQINIGGLAY	1080

QY	1081	PPLPLHEGPRAPSGYSQPPSVCSSTSNFGPFAGGVVSPQPHSSYYSGMTGPOHPFYNR-	1139
Db	1081	PPLPLHEGPRPSPSGYSQPPSVCSSTSNFGPFAGGVVSPQPHSSYYSGLSGQHPFYNR	1140
QY	1140	-----GSGPAPGPVLLL	1151
Db	1141	PFAPYLYTPRYPGGSOHLISRSSVKASLPRDNNGLPCDSGPNKQORGRPCHRQITVT	1200
QY	1152	NSLNVDAVCEKLQIEGLDQSMLEPOYCTTIKKANINGRVLACNIDELKEMMNFGDWH	1211
Db	1201	EFNDRCCRCEKLQIEGLDQSMMPQYCTTIKKANINGRVLSCNIDELKEMMNFGDWH	1260
QY	1212	LFRSTVLEWRNAESHVVPEDPRFLSESSSGPAPHGEPARRASHNELPHTLESSQTYTLN	1271
Db	1261	LFRSMVLEMRSVESQVVPEDPRFLNENSAPVPHGESARRPSHTELPTELSSQTYTLN	1320
QY	1272	FSFEELNTLGLDEGAPRHSNLSQWQOTRTPSLSSLSNSODSSEISKLTQVQAEYRDAY	1331
Db	1321	FSFEELNTLGLDEGAPRHSNLSQWQOTRTPSLSSLSNSODSSEISKLTQVQAEYRDAY	1380
QY	1332	REYIAQMSOLEGGPGSTTISGRSSPHSTYYMGQSSGSIHNSLNEQEKQKQSEPKDDGR	1391
Db	1381	REYIAQMSOLEGGTGSSTISGRSSPHSTYYIGQSSGSIHSTLEQERKEGELKQEDGR	1440
QY	1392	KSFLMKRGDVIDYSSGVSTNDASPLDPITEDEKSDQSGSKLLPGKKSSERSLFTQDL	1451
Db	1441	KSFLMKRGDVIDYSSGVSTNEASPLDPITEDEKSDQSGSKLLPGKKSSERSPLFTQDL	1500
QY	1452	KLKGSGLRYQKLPSDESGTSESDNTPLLKDDKDKAEGKVERVPKSPHSAEPIRTFI	1511
Db	1501	KLKGGGLRYQKLPSDESGTSESDNTPLLKDDKDKAEGKAEVCKSPHSAEPIRTFI	1560
QY	1512	KAKEYLDALLDKQSSDSGVRSSESSPNHSLHNEVADDSQLEKANLILEDDSHSGKRG	1571
Db	1561	KAKEYLDALLDKQSSDSGVRSNESSPNHSLHNEAADDQLEKANLILEDEGHSGKRG	1620
QY	1572	IPHLSGLQDPIIARMSICSEDKKSPSECSLIASSPEENWPACOKAYNLNRTPTVTLLN	1631
Db	1621	MPHLSGLQDPIIARMSICSEDKKSPSECSLIASSPEESWPACOKAYNLNRTPTVTLLN	1680
QY	1632	NSAPANRANQNDMEGIRETSQVILRPSSSNPTTIQENLKSMTHKESQSSSYTRLK	1691
Db	1681	NTAPTNRANQNDMEGIRETSQVILRPSPNPTAVQNLKSMHAKKSQSSSYTRLK	1740
QY	1692	DPPELHAAASSESTGFEERESIL	1715
Db	1741	DASELH-AAASESTGFEERESIL	1763

RESULT 11
AAE25143
ID AAE25143 standard; protein; 1715 AA.
XX
AC AAE25143;
XX
DT 30-OCT-2002 (first entry)
XX
DE Rat ARMS protein.
XX
KW Ankyrin repeat-rich membrane spanning protein; ARMS; neurotrophin;
KW ephrin; receptor tyrosine kinase; growth cone; neuron; neuronal cell,
KW diagnostic; imaging; rat.
OS Rattus norvegicus.
XX
FH Key Location/Qualifiers
FT Region 17..390
FT /note= "Ankyrin repeat"
FT Region 354..493
FT /note= "N-terminal region"
FT Domain 496..518
FT /note= "Transmembrane domain"

Db 1321 DKQAEYDAYREYIAQMSQLEGGTSGSTISGRSSPHSTYYIGOSSGSGSIHSTLQERG 1380
Qy 1381 KDSEPKDDGRKFLMKRGVDIYSSSGVGTNDASPLDPTDEDEKSDQSGSKLLPKGKS 1440
Db 1381 KEGELKQEDGRKFLMKRGVDIYSSSGVGTNEASPLDPTDEDEKSDQSGSKLLPKGKS 1440
Qy 1441 SERSSLFQTLKLGSLRYQKLPSDESGTEESDNTPLKDDKDKRKAEGKVERVPKS 1499
Db 1441 SERPSLFQTLKLGSLRYQKLPSDESGTEESDNTPLKDDKDKRKAEGKVERVPKS 1500
Qy 1500 PEHSABPIRTFIKAKEYLSDALLDKKSDSDGVRSSSESSPNHSLHNEVADDSOLEKANLI 1559
Db 1501 QEHSABPIRTFIKAKEYLSDALLDKKSDSDGVRSSSESSPNHSLHNEVADDSOLEKANLI 1560
Qy 1560 ELDDSHSGRGKTPHLSGLQDPIIARMSTCSDDKSPSCSLIASPENWPAOKAYN 1619
Db 1561 ELEDEHSGRGKTPHLSGLQDPIIARMSTCSDDKSPSCSLIASPENWPAOKAYN 1620
Qy 1620 LNRTPTVTLLNNNSAPANRANQFDEMEGIRETSQVILRPSSSPNPTTIONENLKSMTWK 1679
Db 1621 LNRTPTVTLLNNNSAPANRANQFDEMEGIRETSQVILRPSSSPNPTTIONENLKSMTWK 1680
Qy 1680 RSQRSSYTRLSDKPPPELHAASSESTGFGHERESIL 1715
Db 1681 RSQRSSYTRLSDKPPPELHAASSESTGFGHERESIL 1715

RESULT 12

ADCO6849
ID ADC06849 standard; protein; 1715 AA.
XX
AC ADC06849;
XX
DT 18-DEC-2003 (first entry)
XX
DE Rat Kidins220Pc protein AF313464.
XX
DE cytosolic; prostate cancer; breast; gene therapy; transgenic; rat;
KW Kidins220Pc; kinase D-interacting substrate of 220kDa.
XX
OS Rattus sp.

XX W02003064599-A2.

XX 07-AUG-2003.

XX 24-JAN-2003; 2003WO-US001943.

XX 25-JAN-2002; 2002US-00054935.

XX 14-FEB-2002; 2002US-0356130P.

XX 22-MAR-2002; 2002US-00102946.

XX 08-APR-2002; 2002US-00117229.

XX 14-MAY-2002; 2002US-00144198.

XX 19-JUL-2002; 2002US-00197824.

XX (ORIG-) ORIGENE TECHNOLOGIES INC.

XX Sun Z, Li X, Jay G, Kovacs KF, Fan W, Shu Y;

XX WPI; 2003-679495/64.

XX New isolated polynucleotide related to cancer genes, useful for
PT detecting, diagnosing, staging, monitoring, prognosticating, preventing
PT or treating cancers, e.g. breast and prostate cancers.

XX Disclosure; Fig 20; 128pp; English.

XX The invention relates to a novel isolated polynucleotide comprising a
CC differentially-regulated mammalian cancer gene. The polynucleotides of
CC the invention demonstrate cytostatic activity and are differentially
CC expressed in prostate cancer. The polynucleotide, polypeptides and
CC methods of the invention may be useful for detecting, diagnosing,
CC staging, monitoring, prognosticating, preventing, or treating cancers,

CC particularly breast and prostate cancers. Furthermore, the invention may
CC be utilised during gene therapy procedures or in the production of
CC transgenic animals. The current sequence is that of the prostate cancer-
CC related protein of the invention. The current sequence is not fully
CC legible within the specification; the complete sequence was obtained from
CC GenBank.

XX Sequence 1715 AA;

Qy Query Match 93.0%; Score 8266; DB 7; Length 1715;
Db Best Local Similarity 92.7%; Pred. No. 0;
Matches 1591; Conservative 53; Mismatches 70; Indels 2; Gaps 2;

Qy 1 MSVLISQSVINYVEENIPALKALLEKCKVDNERECGOTPLMIAAEOGNLEIVKELKN 60

Db 1 MSVLISQSVINYVEENIPALKALLEKCKVDNERECGOTPLMIAAEOGNLEIVKELKN 60

Qy 61 GANCNLEDDNWTALISASKEGHVHIVEELLKCGVNLHRDMGGMWTALMWACYKGRDVV 120

Db 61 GANCNLEDDNWTALISASKEGHVHIVEELLKCGVNLHRDMGGMWTALMWACYKGRDVV 120

Qy 121 ELLSHGANPSVTGLQYSVYPIIWAAGRGHADIHLLONGAKVNCSDKYGTTPPLVWAAR 180

Db 121 ELLSHGANPSVTGLQYSVYPIIWAAGRGHADIHLLONGAKVNCSDKYGTTPPLVWAAR 180

Qy 181 KGHLECVKHLAMGADVDOEGANSMTALIVAVKGGYTQSVKTELKKNPNVNLTKDGNTA 240

Db 181 KGHLECVKHLAMGADVDOEGANSMTALIVAVKGGYTQSVKTELKKNPNVNLTKDGNTA 240

Qy 241 LMIASKEGHEIVQDILLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRG 300

Db 241 LMIASKEGHEIVQDILLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRG 300

Qy 301 QDNKTALYWAVEKGNATWVRDILQCNPDTEICTKOGTETPLIKATKRNIEVVELLDKGA 360

Db 301 QDNKTALYWAVEKGNATWVRDILQCNPDTEICTKOGTETPLIKATKRNIEVVELLDKGA 360

Qy 361 KVSANDKKGDTPLHAIARGSRKLAELLARNPKDGLLYRPNKAGETPNYINDCSHOKSIL 420

Db 361 KVSANDKKGDTPLHAIARGSRKLAELLARNPKDGLLYRPNKAGETPNYINDCSHOKSIL 420

Qy 421 TQIFGARHLSPTETDGMGLGYDLYSSALADILSEPTMQPPICVGLYAQWGSFKLLKKL 480

Db 421 TQIFGARHLSPTETDGMGLGYDLYSSALADILSEPTMQPPICVGLYAQWGSFKLLKKL 480

Qy 481 EDEMKTFAQQOIEPLPQPSWLIIVFLTLCCGGIGLLFAFVTHPNLGIASLSFLALIYIF 540

Db 481 EDEMKTFAQQOIEPLPQPSWLIIVFLTLCCGGIGLLFAFVTHPNLGIASLSFLALIYIF 540

Qy 541 FIVIFEGGRREGESMNWAVLSTRLARHIGYLELLKLMFVNPPELPEOTTKALPVRELF 600

Db 541 FIVIFEGGRREGESMNWAVLSTRLARHIGYLELLKLMFVNPPELPEOTTKALPVRELF 600

Qy 601 TDYNRLSSVGGETSIAEMIATLSDACEREFGLATRLFRVFTKEDTQGGKKWKKTCLPS 660

Db 601 TDYNRLSSVGGETSIAEMIATLSDACEREFGLATRLFRVFTKEDTQGGKKWKKTCLPS 660

Qy 661 FVIFLFIICIIIGITLLAIFRVDPKHLTVNAVLSIASVVGFLAVLNCRTWQVLDL 720

Db 661 FVIFLFIICIIIGITLLAIFRVDPKHLTVNAVLSIASVVGFLAVLNCRTWQVLDL 720

Qy 721 NSORKELHNAASKLHKLKSEGFMKVLKCEVELMARWAKTIDSTONOTRLVLIIDGLDAC 780

Db 721 NSORKELHNAASKLHKLKSEGFMKVLKCEVELMARWAKTIDSTONOTRLVLIIDGLDAC 780

Qy 781 EODKVLQMLDTRVRLFSKGFPIAFASDPHIIKAINQNLNSVLRDSNINGHYDMRNIVH 840

Db 781 EODKVLQMLDTRVRLFSKGFPIAFASDPHIIKAINQNLNSVLRDSNINGHYDMRNIVH 840

Qy 841 LPVFLNSRGLSNARKPLVTSATNGDVPCSDTTGIIQBDARRVSQNSIGEMTKLGSKTALN 900

Db 841 LPVFLNSRGLSNARKPLVTSATNGDVPCSDTTGIIQBDARRVSQNSIGEMTKLGSKTALN 900

```
QY 901 RDTYRRQMORTITRQMSFDLTKLLVTEDFMSDISPQTMRLRLNIVSVTGRLLRANQIS 960
Db 901 RDTYRRQMORTITRQMSFDLTKLLVTEDFMSDISPQTMRLRLNIVSVTGRLLRANQIT 960
QY 961 FNDWELASWINLTCQWPTYSWLLIYLBETEGIPQMTLKIYERISKNITPTKDVPELL 1020
Db 961 FNDWELASWINLTCQWPTYSWLLIYLBETEGIPQMTLKIYERISKNITPTKDVPELL 1020
QY 1021 EIDGDIRNFPEFLSRTPLVARDVKVFLPCTVNLDPKRLREIADVRAAREQISIGGLAY 1080
Db 1021 EIDGDIRNFPEFLSRTPLVARDVKVFLPCTVNLDPKRLREIADVRAAREQINIGGLAY 1080
QY 1081 PPLPLHEGPAPSGYSPVSCSTSFNGPAGVSPQPHSSSYSGMTQPHFPYNRG 1140
Db 1081 PPLPLHEGPAPSGYSPVSCSTSFNGPAGVSPQPHSSSYSGMTQPHFPYNRK 1140
QY 1141 SGPARCPVVLNLSNVADVACEKLOIEGLDQSMLOYCTTIKANINGRVLACQNDILK 1200
Db 1141 AVFATGSSLLSSMTVDVCEKLRQIEGLDQNMPOYCTTIKANINGRVLSCQNDILK 1200
QY 1201 KEMNNFQDWHLFRSTVLEMRNAESHVVPEDPRFLSESSSGPAPHCEPARRASHNELPHT 1260
Db 1201 KEMANNFQDWHLFRSWLEMRNESHVESQVVPEDPRFLNENSSAPVPAGESARRSHTLPLT 1260
QY 1261 ELSQTPYTLNPSPEELNTLGLDEGAPRHSNLSWQSTRTTSPSLSSLSQDSSIEISKLT 1320
Db 1261 ELSQTPYTLNPSPEELNTLGLDEGAPRHSNLSWQSTRTTSPSLSSLSQDSSIEISKLT 1320
QY 1321 DKVQBYRDAYREYIAQMSQLEGGPGSTISGRSSPHSTYMGQSSGGSIHSLNLEQKG 1380
Db 1321 DKVQBYRDAYREYIAQMSQLEGGPGSTISGRSSPHSTYMGQSSGGSIHSLNLEQKG 1380
QY 1381 KDSBPDPDGRKSLMKRGDVIDYSSGVSSTNDASPLDPIITEDESKSQSGSKLLPGKKS 1440
Db 1381 KEGELKQEDGRKSLMKRGDVIDYSSGVSSTNEASPLDPIITEDESKSQSGSKLLPGKKS 1440
QY 1441 SERSLFQDLDLKGSLRYOKLPSDEDESGTEESDNTPLKDDKORKAEKQVERVPKS 1499
Db 1441 SERPSLFQDLDLKGGLRYOKLPSDEDESGTGRVQITPHCSKMIRTKRLKAKQRECA 1500
QY 1500 PEHSAPRTTIKAYEYLSDDLKDDSGVRSSESSPHSLNNEVADDSOLEKANLI 1559
Db 1501 QEHSAPRTTIKAYEYLSDDLKDDSGVRSSESSPHSLNNEVADDSOLEKANLI 1560
QY 1560 ELEDHSHGKRGIPHSLSGLODPIIARMSICSEDKKSPSECSLIASSPEENWPAQOKAYN 1619
Db 1561 ELEDHSHGKRGIPHSLSGLODPIIARMSICSEDKKSPSECSLIASSPEESWPAQOKAYN 1620
QY 1620 LNRTPSTVTLNNSAPANRANQNFDEMEGIRETSQVILRPSSSPNPTTIQENLKSMTHK 1679
Db 1621 LNRTPSTVTLNNTAPTNRANQNFDEIEGIRETSQVILRPGSPNPTAVQENLKSMAHK 1680
QY 1680 RSQRSSYRLSKDPPHLLAAASSESTGFEERESIL 1715
Db 1681 RSQRSSYRLSKDSELH-AAASSESTGFEERESIL 1715

RESULT 13
ADC06850
ID ADC06850 standard; protein; 1142 AA.
XX
XX
AC ADC06850;
XX
DT 18-DEC-2003 (first entry)
DE
XX
XX Kidins 220Pc-related protein Pc473.
XX
XX cytosolic; prostate cancer; breast; gene therapy; transgenic;
XX Kidins220Pc; kinase D-interacting substrate of 220KDa.
XX
OS Unidentified.
XX
XX WO2003064599-A2.
PN
```

```
XX 07-AUG-2003.
XX
XX 24-JAN-2003; 2003WO-US001943.
XX
XX 25-JAN-2002; 2002US-00054935.
PR 14-FEB-2002; 2002US-0356130P.
PR 22-MAR-2002; 2002US-00102946.
PR 08-APR-2002; 2002US-00117229.
PR 14-MAY-2002; 2002US-00144198.
PR 19-JUL-2002; 2002US-00197824.
XX
XX (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
XX Sun Z, Li X, Jay G, Kovacs KF, Fan W, Shu Y;
XX
XX WPI; 2003-679495/64.
XX
XX New isolated polynucleotide related to cancer genes, useful for
XX detecting, diagnosing, staging, monitoring, prognosticating, preventing
XX or treating cancers, e.g. breast and prostate cancers.
XX
XX Disclosure; Fig 20; 128pp; English.
XX
XX The invention relates to a novel isolated polynucleotide comprising a
XX differentially-regulated mammalian cancer gene. The polynucleotides of
XX the invention demonstrate cytostatic activity and are differentially
XX expressed in prostate cancer. The polynucleotide, polypeptides and
XX methods of the invention may be useful for detecting, diagnosing,
XX staging, monitoring, prognosticating, preventing or treating cancers,
XX particularly breast and prostate cancers. Furthermore, the invention may
XX be utilised during gene therapy procedures or in the production of
XX transgenic animals. The current sequence is that of the prostate cancer-
XX related protein of the invention. The C-terminus of the current sequence
XX is illegible within the specification and as a result, residues 1143-1715
XX are absent.
XX
XX Sequence 1142 AA;
XX
XX Query Match 66.8%; Score 5931; DB 7; Length 1142;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1142; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSVLISQSVINYVEENIPALKKLEKCKVDNERNECGQTPLMIAAEOGNLEIVKELIKN 60
Db 1 MSVLISQSVINYVEENIPALKKLEKCKVDNERNECGQTPLMIAAEOGNLEIVKELIKN 60
QY 61 GANCNLEDLONWTALISASKEGHVHIIVEELKCGVNLHHRDMGWTALMWACYKGRTDVV 120
Db 61 GANCNLEDLONWTALISASKEGHVHIIVEELKCGVNLHHRDMGWTALMWACYKGRTDVV 120
QY 121 ELLLSHGANSVTLQYQSVYPIIWAAGRGHADIHLLLLONGAKUNCSDKYTTPLVWAAR 180
Db 121 ELLLSHGANSVTLQYQSVYPIIWAAGRGHADIHLLLLONGAKUNCSDKYTTPLVWAAR 180
QY 181 KGHLECVKHLAMGADVDOEGANSMTALIIVAVKGYTQSVKEILKRNPNVNLTKDGN 240
Db 181 KGHLECVKHLAMGADVDOEGANSMTALIIVAVKGYTQSVKEILKRNPNVNLTKDGN 240
QY 241 LMIAESKEGHTEIVQDLDLAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQYADIDIRG 300
Db 241 LMIAESKEGHTEIVQDLDLAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQYADIDIRG 300
QY 301 QDNKTALYWAVEKGNATWVRDILQCNPDTEICTKGTEPLIKATMNRNIEVVELLLDKGA 360
Db 301 QDNKTALYWAVEKGNATWVRDILQCNPDTEICTKGTEPLIKATMNRNIEVVELLLDKGA 360
QY 361 KVSADVKKGDTPLHIAIRGRSRKLAELLRNPKDGRLLYRPNKAGETPNYDCSHQSKIL 420
Db 361 KVSADVKKGDTPLHIAIRGRSRKLAELLRNPKDGRLLYRPNKAGETPNYDCSHQSKIL 420
QY 421 TQIFGARHLSPTETDGMGLGYDLYSSALADILSEPTMQPPICVGLYAGWGSKGFLLK 480
Db 421 TQIFGARHLSPTETDGMGLGYDLYSSALADILSEPTMQPPICVGLYAGWGSKGFLLK 480
```

```
Db 421 TQIFGARHLSPTETDGMGLDYSSALADILSEPTMQPPICVGLYAQMGSFKLLKKL 480
Qy 481 EDEMKTFAGQOIEPLPQFQSWLIIVFLTLCCGGLLFAFTVHNLGIAVSLSFALLIYIF 540
Db 481 EDEMKTFAGQOIEPLPQFQSWLIIVFLTLCCGGLLFAFTVHNLGIAVSLSFALLIYIF 540
Qy 541 FIVYFGRRGEGESWNAWLSRLARHIGLYLELLKLMFVNPPPELPEQTTKALPVRFLF 600
Db 541 FIVYFGRRGEGESWNAWLSRLARHIGLYLELLKLMFVNPPPELPEQTTKALPVRFLF 600
Qy 601 TDYRLSSVGGTSLAEMIATLSDACEREFGLATRLFRVFKTEDTGKKKWKTCCLPS 660
Db 601 TDYRLSSVGGTSLAEMIATLSDACEREFGLATRLFRVFKTEDTGKKKWKTCCLPS 660
Qy 661 FVFLFPIIGCIISGITLLAIFRVDPKHLTVNAVLIASVVGAFVNLNCRTHWQVLDL 720
Db 661 FVFLFPIIGCIISGITLLAIFRVDPKHLTVNAVLIASVVGAFVNLNCRTHWQVLDL 720
Qy 721 NSOKRRLHNAASKLHLKSGFMKVLKCEVELMARMAKTIDSTFQNTQRLVWIIDGLDAC 780
Db 721 NSOKRRLHNAASKLHLKSGFMKVLKCEVELMARMAKTIDSTFQNTQRLVWIIDGLDAC 780
Qy 781 EQDKVLQMLDTRVLFSGKPFIAIFASDPHIIKAINQNLSVLRDSNININGHDYMRNIVH 840
Db 781 EQDKVLQMLDTRVLFSGKPFIAIFASDPHIIKAINQNLSVLRDSNININGHDYMRNIVH 840
Qy 841 LPVFLNSRGLSNARKFLVTSATNGDVPDSCDTTGIQEDADRRVQNSLIGEMTKLGSKTALN 900
Db 841 LPVFLNSRGLSNARKFLVTSATNGDVPDSCDTTGIQEDADRRVQNSLIGEMTKLGSKTALN 900
Qy 901 RRDYTRRRQMTITROMSFDLTKLVTEDWFSDISPOTMRRLLNIVSVTGRLLRANOIS 960
Db 901 RRDYTRRRQMTITROMSFDLTKLVTEDWFSDISPOTMRRLLNIVSVTGRLLRANOIS 960
Qy 961 FNMRLASWINLQEQPYRTSWLILYEETEGIPDQMTLXIYIERISKNIPTTKDVEPLL 1020
Db 961 FNMRLASWINLQEQPYRTSWLILYEETEGIPDQMTLXIYIERISKNIPTTKDVEPLL 1020
Qy 1021 EIDGDIRNFVFLSSRTPVLVARDVKVFLCTVNLDPKLEIITADVRAAREQISIGGLAY 1080
Db 1021 EIDGDIRNFVFLSSRTPVLVARDVKVFLCTVNLDPKLEIITADVRAAREQISIGGLAY 1080
Qy 1081 PPLPLHEGPPRAPSGVSOQPSVCSSTSFNGPFGVVSPOPHSSYSGMTGPQHPFYNRG 1140
Db 1081 PPLPLHEGPPRAPSGVSOQPSVCSSTSFNGPFGVVSPOPHSSYSGMTGPQHPFYNRG 1140
Qy 1141 SG 1142
Db 1141 SG 1142

RESULT 14
ADCO6846
ID ADCO6846 standard; protein; 1184 AA.
XX
AC ADCO6846;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human Kidins220Pc protein XM_045362.
XX
KW cytostatic; prostate cancer; breast; gene therapy; transgenic; human;
KW Kidins220Pc; kinase D-interacting substrate of 220kDa; chromosome 2p25.1.
XX
OS Homo sapiens.
XX
PN W02003064599-A2.
XX
PD 07-AUG-2003.
XX
PF 24-JAN-2003; 2003WO-US001943.
XX
PR 25-JAN-2002; 2002US-00054935.
```

```
PR 14-FEB-2002; 2002US-0356130P.
PR 22-MAR-2002; 2002US-00102946.
PR 08-APR-2002; 2002US-00117229.
PR 14-MAY-2002; 2002US-00144198.
PR 19-JUL-2002; 2002US-00197824.
XX
PA (ORIG-) ORIGENE TECHNOLOGIES INC.
XX
PI Sun Z, Li X, Jay G, Kovacs KF, Fan W, Shu Y;
XX
XX MPI; 2003-679495/64.
XX
PT New isolated polynucleotide related to cancer genes, useful for
PT detecting, diagnosing, staging, monitoring, prognosticating, preventing
PT or treating cancers, e.g. breast and prostate cancers.
XX
PS Disclosure; Fig 20; 128pp; English.
XX
CC The invention relates to a novel isolated polynucleotide comprising a
CC differentially-regulated mammalian cancer gene. The polynucleotides of
CC the invention demonstrate cytostatic activity and are differentially
CC expressed in prostate cancer. The polynucleotide, polypeptides and
CC methods of the invention may be useful for detecting, diagnosing,
CC staging, monitoring, prognosticating, preventing or treating cancers,
CC particularly breast and prostate cancers. Furthermore, the invention may
CC be utilised during gene therapy procedures or in the production of
CC transgenic animals. The current sequence is that of the prostate cancer-
CC related protein of the invention. The current sequence is not fully
CC legible within the specification; the complete sequence was obtained from
CC GenBank.
XX
SQ Sequence 1184 AA;
```

```
Query Match 66.4%; Score 5899.5; DB 7; Length 1184;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1138; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 MSVLISQSVINYYEENIPALKALLECKDVERNECGOTPLMIAAEOGNLEIVKELIKN 60
Db 1 MSVLISQSVINYYEENIPALKALLECKDVERNECGOTPLMIAAEOGNLEIVKELIKN 60
Qy 61 GANCNLEDLNDNTALISASKEGHVHIVEELLKCGVNLHHRDMGWTALMWACYKGRDVV 120
Db 61 GANCNLEDLNDNTALISASKEGHVHIVEELLKCGVNLHHRDMGWTALMWACYKGRDVV 120
Qy 121 ELLLSHGANSPTGLQYSYVPIIWAAGRCHADIVHLLLONGAKVNCSDKYGTTPLVWAAR 180
Db 121 ELLLSHGANSPTGLQYSYVPIIWAAGRCHADIVHLLLONGAKVNCSDKYGTTPLVWAAR 180
Qy 181 KGHLECVKHLAMGADVDOEGANSMTALIVAVKGGYTQSVKEILKRNPNVNLTKDGNTA 240
Db 181 KGHLECVKHLAMGADVDOEGANSMTALIVAVKGGYTQSVKEILKRNPNVNLTKDGNTA 240
Qy 241 LMIASKEGHEIIVQDILLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIRG 300
Db 241 LMIASKEGHEIIVQDILLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIRG 300
Qy 301 QNKKTALYWAVEKGNATWVRDILQCNPDTEICTKGETPLIKATKVRNIEVVELLDKGA 360
Db 301 QNKKTALYWAVEKGNATWVRDILQCNPDTEICTKGETPLIKATKVRNIEVVELLDKGA 360
Qy 361 KUSAVDKKGDTPHIAIRGRSKLAELLARNPKDGLLRPNKAGETPNYIDCSHOKSIL 420
Db 361 KUSAVDKKGDTPHIAIRGRSKLAELLARNPKDGLLRPNKAGETPNYIDCSHOKSIL 420
Qy 421 TQIFGARHLSPTETDGMGLDYSSALADILSEPTMQPPICVGLYAQMGSFKLLKKL 480
Db 421 TQIFGARHLSPTETDGMGLDYSSALADILSEPTMQPPICVGLYAQMGSFKLLKKL 480
Qy 481 EDEMKTFAGQOIEPLPQFQSWLIIVFLTLCCGGLLFAFTVHNLGIAVSLSFALLIYIF 540
Db 481 EDEMKTFAGQOIEPLPQFQSWLIIVFLTLCCGGLLFAFTVHNLGIAVSLSFALLIYIF 540
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2005, 19:18:20 ; Search time 40.5 Seconds
(without alignments)
3161.065 Million cell updates/sec

Title: US-10-021-571-4

Perfect score: 8884

Sequence: 1 MSVLISQSVINVEENIPA.....LHAAASSESTGFEERESIL 1715

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep:*
- 6: /cgn2_6/ptodata/1/iaa/backfiles.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	557.5	6.3	1883	4	US-09-949-016-9010
2	557.5	6.3	1883	4	US-09-949-016-9011
3	557.5	6.3	1883	4	US-09-949-016-9012
4	557.5	6.3	1883	4	US-09-949-016-9013
5	557.5	6.3	1883	4	US-09-949-016-9014
6	557.5	6.3	1883	4	US-09-949-016-9015
7	557.5	6.3	1883	4	US-09-949-016-9016
8	557.5	6.3	1883	4	US-09-949-016-9017
9	555.5	6.3	1856	4	US-09-949-016-6964
10	555.5	6.3	1880	4	US-09-949-016-5876
11	555.5	6.3	1881	4	US-09-949-016-6965
12	549	6.2	2753	4	US-09-949-016-7659
13	549	6.2	2753	4	US-09-949-016-7660
14	539	6.1	3924	4	US-09-538-092-1246
15	530.5	6.0	1839	2	US-09-172-977-4
16	530.5	6.0	1839	4	US-09-404-108-4
17	522	5.9	1719	4	US-09-949-016-6966
18	520	5.9	4377	4	US-09-949-016-6978
19	493.5	5.6	843	2	US-09-172-977-3
20	493.5	5.6	843	4	US-09-404-108-3
21	480.5	5.4	1745	2	US-09-031-485-33
22	480.5	5.4	1745	2	US-08-847-429A-33
23	480.5	5.4	1745	3	US-09-065-474-33
24	480.5	5.4	1745	3	US-09-557-034-33
25	444	5.0	994	4	US-10-164-595-38
26	443.5	5.0	3913	4	US-09-949-016-10933
27	438	4.9	1088	3	US-09-082-059-2

28	437	4.9	1881	4	US-09-949-016-8307	Sequence 8307, Ap
29	409.5	4.6	683	4	US-10-164-595-71	Sequence 71, Appl
30	409.5	4.6	1423	3	US-08-810-712-10	Sequence 10, Appl
31	409.5	4.6	1431	4	US-09-538-092-1198	Sequence 1198, Ap
32	384.5	4.3	1704	4	US-09-392-812A-2	Sequence 2, Appl
33	376.5	4.2	1327	3	US-09-196-387-2	Sequence 2, Appl
34	376.5	4.2	1327	4	US-09-841-835-2	Sequence 2, Appl
35	376.5	4.2	1327	4	US-09-972-115A-8	Sequence 8, Appl
36	376.5	4.2	1619	4	US-09-392-812A-4	Sequence 4, Appl
37	374	4.2	348	2	US-09-031-485-28	Sequence 28, Appl
38	374	4.2	348	2	US-08-847-429A-28	Sequence 28, Appl
39	374	4.2	348	3	US-09-065-474-28	Sequence 28, Appl
40	374	4.2	348	3	US-09-557-034-28	Sequence 28, Appl
41	373.5	4.2	546	4	US-09-949-016-7169	Sequence 7169, Ap
42	370.5	4.2	949	3	US-09-196-387-10	Sequence 10, Appl
43	370.5	4.2	949	4	US-09-841-835-10	Sequence 10, Appl
44	370	4.2	1166	4	US-09-972-115A-6	Sequence 6, Appl
45	370	4.2	1227	4	US-09-849-602-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1

US-09-949-016-9010
; Sequence 9010, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9010
; LENGTH: 1883
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9010

Query Match 6.3%; Score 557.5; DB 4; Length 1883;
Best Local Similarity 19.7%; Pred. No. 1.9e-38;
Matches 374; Conservative 286; Mismatches 659; Indels 581; Gaps 72;

Qy	21	LKALLEKCDVDERNECGQTFPLMAEQGNLEIVKELIKNGANCNLEDLDNWTALISASK	80
Db	97	VRELNVYGANVNAQSQKGFPLYNAAQGNHLEVVVFLLENGANQNVATEDGFTFLVALQ	156
Qy	81	EGHVHIVEELLKCGV-----NLEHRDMGCGTALMWA	111
Db	157	QGHENVVAHLINVTGKGVLPALHIAARNDDTFTAALLQNDPNPVLSTGFTPLHIA	216
Qy	112	CYKGRDQVVELLSHGANSPT-----GLQY-----SV	139
Db	217	AHVNLNVAQLLNRRGASVNFPTQGITPLHIASSRRGNVIMVRLLDORGAIETKTDEL	276
Qy	140	YPIWAAGRADIIVHLLONGAKVNSDKYGTITPLVAAKGHLECVKHLANGADV-	198
Db	277	TPHCAARNHVRISIELLDHGAPIOAKTNGLSPIHMAAQGDHLDVRLQLQDAEIDD	336
Qy	199	-----ORGA-----NSMTALIVAVKGGYTSQVKELKR	226
Db	337	ITLDHLTPHVAHCHGHRVAKVULLDKGAKNSPALNGFTPLHIAKKNHVRVWELLKT	396
Qy	227	NPVNLTKDQNTALMIASKEGHEIYQDILLDAGTYVNIIPDRSGDTVLIGAVRGHVEIV	286

Db 397 GASIDAVTESGLTPLHVASPMGHLPIVKNLQKASPNVSNVKTPLHMAARAGTEVA 456
QY 287 RALLQYADIDIRQDNKTALYWAKEGNATWVRDILQCNPDTEICTKDGETPLIKATKM 346
Db 457 KYLLQNKAKVNAKDDQTPHCAARIGHTNMVKKLLLENNANFNLTAGHTPLHIAARE 516
QY 347 RNTVEVELLDKAKVASDKGDTPLHIAIORSRKLAEALLLNRNPKDGRLLYPRNKAGE 406
Db 517 GHVETVLLALLEKASOACTKGTGFTPLHVAAKYGVKRVAEELLE--RDA----HPNAAGK 570
QY 407 ---TPYNDICSHQKSLTQIFGARHLSPTECDGMLGY-DLYSSALADILSEPTMOPPCC 462
Db 571 NGUTPLHVAHNHNLIDVKKLLPRGSGSPAWN--GYTPPLHIAKQN-----QVEVA 621
QY 463 VGLYAQWGSKSPKLLKLEDEMKTFAQQTEPIQFQSWLIVFTLLLCG-----LGLLFA 518
Db 622 RSLQYGGSAARESV-----QGVTPH-----LAAQGHAEWVALLLS 659
QY 519 FTVHPNLGIAVSFLALLYIFVIYFGRRREGESNMAWVLSRLARHIGVLELLK-- 577
Db 660 KOANGNLGNKSGLTPLHLV-----AQBG-----HVPVADVLIXH 693
QY 578 -----LMFVNPPPELPEQTKALPVRELF--TDYNRLLSVG-----GETSLA 616
Db 694 GNMVDATTRMGYTPLVASHYGNIKLVFLLOHQADVNAKTKLGYSPHLQHAQQGHTDIV 753
QY 617 EMI-----ATLSDACEREFGLA--TRLFRVFKTEDTOGKKWKTKCCLPSPFV 662
Db 754 TLLKNGASPNVSSDGTPLATAKRLGIYSVTDVLKVV-TDET-----SFV 799
QY 663 IFLFIIGIISGITLAIARFVDPKHL-----TNAVLIASIVG---LAFVLCNRTWQ 714
Db 800 L-----VSDKHRMSPFETVDEIL-DVSEDEBELISFKAERR-----835
QY 715 VLDSLNSQKRLHNAASKLHLKSEGFMKVLCVELMARMAKTIDSFTQNTRLVVII 774
Db 836 --DSRDVDEKEILDVPKLDQVVESPAIPRIPE-----AMPETVIRSEEGQASKEY 897
QY 775 DGLDACEQDKVLQMLDTRVRLFS--KGFPIAIPASDP-----HIK-----814
Db 888 DEDSLPSPATSDNISPAVSPVHTGFLVSPWDARGSMGRSHRGLRVVIPRTCA 947
QY 815 -----ANQNLNSVLRDSNINGHDYMN--IVHLPVFL-NRGLS 851
Db 948 APTRITRLVKPQKLSPTPPPLAEEGLASRIIALGPTGAQFLSPVIVEIPHAFSHGRG-- 1005
QY 852 NARKFLVTSATNGDVPCSDTTGQEDADRVSONSIGEMTKLGSKTALNRDTRRRQWQ 911
Db 1006 -DRELVLVRSNGSV-----WKEHRSRYGESYLDQILN-GMDEBELGSLELEKRVCC 1055
QY 912 RTITRQMSFOLTLLVLT---EDWFSDISPTQMRLLNIVSVTGRLLRANQISFNWDLAS 968
Db 1056 RIIT--TDPLFYVIMSRLCQD-YDYGPEG-----GSLKSLVPLVQATPENAVTK 1105
QY 969 WINLTQWPYRWSWLIYLEETEGIPQWTLK-----TIYERISK-----1008
Db 1106 RVKLAQ-----AQVPDELVTKLGNQATPSPIVTVPRRRKFRPIGLR 1151
QY 1009 -NIPTTKQVPELLEIDGDIRNFEV-----LSSRTPVLVARDVKVFLPCTV 1053
Db 1152 IPLPPSWTDNPRDSGEGDTLSRLLCVSGIGTDOAQWEDITGTYKLYANECANF---TT 1208
QY 1054 NLPDKLREIADYRAAREQISIGLAYPPLPLHEGPPRAPSGYSPPSVCSSTSFNGPFA 1113
Db 1209 NV--SARFWSLDCPRTAEAVNPATLLYKEL-----TAPYMAKFV 1246
QY 1114 GGVSVPQSHSYSGMTGPOHPFYNRGSGPAPGV--VLLNSLNVDAVCE-----K 1162
Db 1247 -----IFAKMNDPREGLRCYCMTDDKVDKTLFQEHENFVEVAR 1284
QY 1163 LKQIEGLDQSWLPQYCTTIKKANINGVLAQCNIDELKEMNWFQDHLFRSTVLEMRN 1222

Db 1285 SRDIEVLECMSL-----FAELSG-----NLVPVKKAAQORSFHFQSFRENRLAM-- 1328
QY 1223 AESHVVPEDPRFLSSSSSGPAPGPEPARASHNE-----LPHTELS----- 1263
Db 1329 -----PVKVRDSSRPFPGSLSFLRKAMKYEDTQHILCHLNITMPPCAKSGAEDRR 1379
QY 1264 -SQTPTLNFPEELNTLGLDEGAPR-----HSNLSWOSQTRTP-SLSSLSNQD 1311
Db 1380 RPTPLALRYSILSSSTPGSLSGTQOAKMMMAVISEHLGLSWAELARELQFVEDIN--- 1436
QY 1312 SSIEISKLTQVQAEYRDAYREYIAQMSQL-----EGGPGSTTISGRSSPHSTYIMQSSS 1367
Db 1437 -----RIEVENPNSLLEQSVALLNLWIREG-----QVANMENLYTALQSID 1478
QY 1368 GSGIHSNLFQEKQKQSEPKDDGRKSFMLKRGVDVITYSSGVSSTNDASPLDPIEDE-- 1425
Db 1479 RGEIVNMLEGSRQSRNLKPD-----RRHTDRDYSLSPSQMGNGYSL-----QDELL 1525
QY 1426 -----KSDOSGSKL-----LPGKSSERSSLFQTDLKLKSGSLRYQKLPSDE 1467
Db 1526 SPASLGALSPLRADQYWNVAVLDAIPLAATEHDTMLESDMQVWSAGLTPSLVTAED 1585
QY 1468 DE---SGTESDNTPLKDDKDKAEKGVKVERVPKSPHSAEPIRTFIKAKEYLSDALLDK 1524
Db 1586 SSLECSKAEDSAT-----GHEWKLEGALSERPGPE-----LGSLELVEDDVTDS 1631
QY 1525 KDSS-----DSGVRSSESPNHLNHEVADDQLEKANLIELEDDSHSGKRGPHSL 1576
Db 1632 DATNGLIDLLEQEBEQSRSEKLPGSKRQDDATGAGQ-DSENEVSLVSGHQRGQARITHS- 1689
QY 1577 SGLQDPIIARMSICSDEK-----KSPSECSLIASSPEENW 1611
Db 1690 -----PTVSQVTERSQRLQDWDADGSIVSYLQDAAQGSW 1724

RESULT 2

US-09-949-016-9011
; Sequence 9011, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9011
; LENGTH: 1883
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9011

Query Match 6.3%; Score 557.5; DB 4; Length 1883;
Best Local Similarity 19.7%; Pred. No. 1.9e-38;
Matches 374; Conservative 286; Mismatches 659; Indels 581; Gaps 72;

QY 21 LKALKECKDVBNERCGOTPLMIAAEOGNLEIVKELIKNGANCNLEDNMTALISASK 80
Db 97 VRELNVYGANVNAQSQKGTPLYMAAQENHLEVVKELLENGANQNVATEDGFTPLVALQ 156
QY 81 EGHVHVIEELLKGV-----NLEHRDMGOWTALMWA 111
Db 157 QGHENVVAHLINVTGKVRPLPALHIAARNDDTRTAAVLLQNDPNPDVLSKTGFTPLHIA 216
QY 112 CYKGRDVTVELLSHGANSVT-----GLQY-----SV 139


```
; Sequence 9014, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9014
; LENGTH: 1883
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-9014

Query Match      6.3%; Score 557.5; DB 4; Length 1883;
Best Local Similarity 19.7%; Pred. No. 1.9e-38;
Matches 374; Conservative 286; Mismatches 659; Indels 581; Gaps 72;

QY 21 LKALLEKCDVDNERCGTPTLMIAAEQGNLEIVKELIKNGANCNLEDLDNWTALISASK 80
DB 97 VRELNVYGANVAQSQKGFPLYMAAQENHLEVVKELLENGANQNVATEDGFTPLAVALQ 156
QY 81 EGHVHIVEELLKGV-----NLEHRDMGWTALMWA 111
DB 157 QGHENVVAHLINTGKGVKRLPALHIAARNDTRTAVALQNDPNPDVLSKTGFTPLHIA 216
QY 112 CYKGRDVTVELLSHGANSV-----GLOY-----SV 139
DB 217 AHYENLVAQLLNRGASVNTFQNGITPLHIAARRGNVIMVRLLLDRGAQIEYTKDEL 276
QY 140 YPTIWAAGRHADIVHLLQNGAKVNSDKYGTTPVMAARKHLCVXKHLAMGADV- 198
DB 277 TPLCAARNGHVRISIELLDHGAPIOAKTKNGLSPIHMAAQGDHLLDCVRLLLQYDAEIDD 336
QY 199 -----QEGA-----NSMTALIVAKGGYTSQVKELKR 226
DB 337 ITLDHLTPLVAAHCGHRRVAKVLLDKGAKPNRSLNGFTPLHIACKONHVRVWELLKT 396
QY 227 NPNVNLTDKDGNTALMASKEGHEIIVQDLDAGTYVNIIPDRGDTVLIGAVRGHVEIV 286
DB 397 GASIDAVTESGLTPLHVASFWGHLPIVKNLLQRGASPNVSNVKVETPLHMAARAGTEVA 456
QY 287 RALLQKYADIDIRQDNKTALYAWVEKGNATWVRDILQCNPDTEICTKGETPLIKATKM 346
DB 457 KYLLQNKAKVNAKADQDTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTPLHIAARE 516
QY 347 RNTVEVELLDKAKVSAVDKGDTPHIAIRGSRKLAELLILRNPKDGLLRYRNKAGE 406
DB 517 GHVETVLALLEKASQACMTKGFPLHVAAKYGVKRVASLELLE--RDA----HPNAAGK 570
QY 407 ---TPVNICDSHQKSLTQIFGARHLSPTETDGMGLY-DLYSALADILSEPTMOPPIC 462
DB 571 NGITPLHVAHNNLDIVKLLPRGSGSPHSPAWN--GYTPLHIAAKQN-----QVEVA 621
QY 463 VGLYAQWGSKSLFKLEDEMKTFAQQIETPLFQFSWLIVFTLLLCGG-----LGLLFA 518
DB 622 RSLQYGGSANAEV-----QGVTPH-----LAAQEGHAEMVALLS 659
QY 519 FTVHPNLGIAVSLFALLIYPIFIVYFGRGREGSNWAWVLSRLARHIGVLELLK- 577
DB 660 KOANGNLGNKSLGTLPLV-----AQEG-----HVPVADVLIKH 693
QY 578 -----LMFVNPPELPEQTIKALPVRFPLF-----TDYNRLSSVY-----GETSLA 616
```

694 GVMVDATTBMGYTPLHVASHYGNIKLVKELLQHOADVNAKTKLGVSPHQAQQOCHTDIV 753
617 EMI-----ATLSDACERBEFGFLA-TRLPFRVFKTETDQKKKWKTKCCLPSFV 662
754 TLLKNGASPNVSSDGTTPLAIAKRLGVISVTDVLKVY-TDET-----SFV 799
663 IFLFIIGCIISGITLLAIIPRVPKHL-----TNAVLISIASVVG--LAFVLNCRWTWQ 714
800 L-----VSDKHMSPFETVDEIL-DVSEDEGELSFAKARR-----835
715 VLDSLINSQRKLNHAASKHLKSEGMVKLVKCEVELMARMARKTIDSTQNTQRLVVI 774
836 --DSRDVDEKELLDPVKLDQWVESPAIPRIPC-----AMPETVIRSEEQOASKEY 887
775 DGLDACEQKVQLMDTVRVLFS--KGPFIAPASDP-----HIILK-----814
888 DEDSLIPSSPATETSDNISPVASPVHTGFLVFMVDARGSMRGRSHNGLRVVIPPRTCA 947
815 -----AINQNLNSVLARDSNINGHYDMRN-IVHLPVFL-NSRGLS 851
948 APTRITCRVLVKPKLSTPPPLAEEGLASRIIATGPTGAQFLSPVIVEIPHAFASHGRG-- 1005
852 NARKFLVTSATNGDVPSCDITGIEDADRRVSNLSGENTKLGSKTALNRRDITYRRQMQ 911
1006 -DRELVLRSSEGSV-----WKEHRSRYGESYLDQILN-GMDEELGSLEELEKRVK 1055
912 RIITROMSFDLTKLVT---EDWFSDISQTMRRLLNIVSVTGRLLRANQISFNWDLAS 968
1056 RIIT--TDFPLFYVMSRLCQD-YDTIGPEG-----GSLSKSLVPLVQATPENAVTK 1105
969 WINLITQMPYRTSWLLIYLEETEGIPDOMTLK-----TIVERISK-----1008
1106 RVKLAQ-----AQVPDELVTKLGNQATFSPVITVEPRRKRHRPIGLR 1151
1009 -NIPTTKVEPLEIDGDIRNFEVF-----LSRTPVLVARDVKVFLPCTV 1053
1152 IPLPSSWTDNPRDSGEGDTLSRLLCVIGGTDOAQWEDITGTTKLVIYANECANF---TT 1208
1054 NLDPKLEIADVRAAREQISIGGLAYPPLPHLEGPPRAPSQVSPSCSSTSFNGPFA 1113
1209 NV--SARFWSLDCPRTAEAVNFATLLYKEL-----TAVPYMAKFV 1246
1114 GGVSQPHSSVYSGMTGPHFPYNRGSGPAGPV--VLLNSLNVDAVCE-----K 1162
1247 -----IFAKNDPREGRLRCYCMTDKDKVKTLEOHENFVEVAR 1284
1163 LKQIEGLDQSMPLPYCTTIKKANINGRVLQACNIDELKKEMNNMFGDWHLFRSTVLEMRN 1222
1285 SRDIEVLEGMSL-----FAELSG-----NLVPVKAAQORSFHFQSFRENRLAM-- 1328
1223 AESHVVPDPRFLSESSSGPAPHGEPARRASHNE-----LPHTELS-----1263
1329 -----PVKVRDSREPGSLSFLRKAMKYEDTQHLCHLNITMPPCAKSGGAEDRR 1379
1264 -SOTPYTLNFSPEELNLTGLDEGAPR-----HSLNSWQSOTRRTP-SLSSLSNSQD 1311
1380 RPTPLALRYSLSESTPGSLSGTEQAEMKMAVISEHLGLSWAELARELQFVSVDIN--- 1436
1312 SSIETSKLTKVQAEYRDAYREYIAQMSQI-----EGPGSTTISGRSSPHSTTYMQOSS 1367
1437 -----RIRVENPNSLLEQSVALNLNWIWREG-----QANMENLYTALQSID 1478
1368 GGSIHNSHLBQEKDSEPKPDGDKRKFMLKRGDVIDYSSGVSTNDASPLDPITEDE-- 1425
1479 RGEIVNMLEGSGRQSNLKP-----RRHTDRDYSLSPSQMNQYSSSL-----QDELL 1525
1426 -----KSDQSGSKL-----LPGKKSSESSSLFQTDKLGKSLRKYQLKPSDE 1467
1526 SPASLCALSSPLRADQYNEVAVLDAIPLAATEHTMTLEMSDMQVWAGLTSLVTAED 1585
1468 DE---SGTBESONTPLLKDDKORKAEGKVERVPKSPHSAEPIRTTIFIKAEYLSDALDK 1524
1586 SSLECSKAEDSDAT-----GHEWKLEGALSSEPRGPE-----LGSLELVEDDITVDS 1631

Qy 1525 KDSS-----DSGVRSESSPNHSLHNEVADDSQLEKANLIBLEDSSHSGKRGIPHS 1576
Db 1632 DATNGLDLLEQEBQSEELPGSKRQDDATGAGQ-DSENEVSLVSGHQRGQARIITHS- 1689
Qy 1577 SGLQDPIIARMSICSEDK-----KSPBECSLIASPBNW 1611
Db 1690 -----PTVSQVTERSQDLQMDADGSIYSYLODAAQGSW 1724

RESULT 6
US-09-949-016-9015
; Sequence 9015, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYNORPHIS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9015
; LENGTH: 1883
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9015

Query Match

Best Local Similarity 19.7%; Pred. No. 1.9e-38;

Matches 374; Conservative 286; Mismatches 659; Indels 581; Gaps 72;

Qy 21 LKALLECKVDRENECGTPLMTAAEQGNLEIVKELIKNGACNLLEDLNTWALISASK 80
Db 97 VRELNVGANVNAQSGFTPLYNAAQENHLEVVKFLLENGANQVATEDGFTPLVALQ 156
Qy 81 EGHVHIVEELKCGV-----NLBHRDGGWTAALMWA 111
Db 157 QGHENVVAHLINYGCKVRLPALHIAARNDDTHTAAVLLQNDPNPDLVSKTGTPLHIA 216
Qy 112 CYKGRDVELLSHGANSV-----GLQY-----SV 139
Db 217 AHVENLVAQILLNRGASVNFPTQNGITPLHIASRRGNVIMVRLLLDRGAQIETKTKDEL 276
Qy 140 YPIIWAAGRHADIVHLLQNGAKVNCSDKYGTTPLVWAARKHLECVKHLAMGADV- 198
Db 277 TPLHCAARNGHVRISEILLDHGAPIQAKTNGLSPIHMAAQGDHLDVRLLLQVDAEIDD 336
Qy 199 -----ORGA-----NSMTALIVAVKGGYVTSVKELKR 226
Db 337 ITLDHLTPLVAAHCRRHVAKVLLDKGAKPNSBALNGFTPLHIACKKNHVRVWELLKLT 396
Qy 227 NPNVNLTKDQNTALMTASKEGHTIEIVODLLDAGTYNIPDRSGDTVLI GAVRGHVEIV 286
Db 397 GASIDAVTESGLTPLHVASFMGHLPIVKNLQRGASPNVSNVKTETPLHMAARAGHTEVA 456
Qy 287 RALLOKQADIDIRQDNKTALYAVEGNATWYRDILQCPDTEICTKDGETPLIKATM 346
Db 457 KYLLQNAKAVNAKADDTPLHCAARIGHTNMVKKLENNANPNLATAGTTPHIAARE 516
Qy 347 RNTVEWELLDKGAKVAVDKGTPLHIAIRGRSRKLAELLNPKDGRLLVPRPKAGE 406
Db 517 GHVETVLALLEKEASQACTYKGTPLHVAKYKRVVAELLE--RDA-----HPNAGK 570
Qy 407 ---TPYINDCSHQSIILTQIFGARHLSPTETDGMGLY-DLYSSALADILSEPTMQPPIC 462
Db 437 -----RIRVENPNSLLEQSVALLNLVWIREG-----QANMENLYTALQSID 1478

Db 571 NGLTPLHVAVHNHNDIVKLLPRGSGPHSPAWN--GYTPLHIAAKQN-----QVEVA 621
Qy 463 VGLVAQWQSGKSKFLKKLEDEMKTFAQQIETPLFPQSWLIVFLTLKCGG---LGLLFA 518
Db 622 RSLLOYGGSNAESV-----QGVTPH-----LAQEGHAEMWALLLS 659
Qy 519 FTVHPNLGIAVSLFALLYIFVIYFGRRREGESWNWAVLSTRLARHIGYLELLK- 577
Db 660 KOANGNLGNKSGLTPLHLV-----AQEG-----HVPVADVLKIH 693
Qy 578 -----LMFVNPPELPEQTTKALPVRPLP--TDYNRLLSSVG-----GETSLA 616
Db 694 GVMVDATTRMGYTPHVAHYGNIKLVKFLLOHADVNAKTKGLGYSPLHQAQOQHTDIV 753
Qy 617 EMI-----ATLSDACEREFGLA--TRLPFRVFKTDTGKKWKKTCCLPSEFV 662
Db 754 TLLKNGASPNVSSDGTTPLATAKRLGYISVTDVLKV--TDEI-----SFV 799
Qy 663 IFLFIIGCIISGITLLAIFRVDPKHL-----TVAVALISIASVVG---LAPVLNCRTWQ 714
Db 800 L-----VSDKHRMSFPETVDEIL--DVSEDEGEELISPKAERR--- 835
Qy 715 VLDSLNSQKRLHNAASKLHLKSEGFMKVLKCEVELMARMAKTIDSTFQNTQRLVVII 774
Db 836 --DSRDVDEEKELDFVPKLDQVVESPAIPRIPC-----AMPETVIVIRSEQOASKEY 887
Qy 775 DGLDACQDQVLQMLDTRVLFSS--KGPFTAIASDP-----HIITK----- 814
Db 888 DEDSLIPSPATETSDNISPVASPVHTGFLVSMVMDARGGSMGRSRLNGRLVPIPRPTCA 947
Qy 815 -----AINQNLSVLDRSININGHYMRN--IVHLPVFL--NSRGLS 851
Db 948 APTRITCRLVKPKLSTPPPLAEEGLASRIIALGPTGAQFLSPVIVEIPHAFSHGRG-- 1005
Qy 852 NAKKFLVTSATNGDVPCSDTTGQEDADRRVQNSLGEMTKLGSKTALNRRDTRRRQMQ 911
Db 1006 -DRELVLRSENGSV-----WKEHRSRYGESYLDQILN--GMDEELGSLLEEKKRVC 1055
Qy 912 RTITRQMSPLTKLLVT---EDWFSDISPTMRRLNIVSVTGRLLRANQISFNWDRLAS 968
Db 1056 RIIT--TDFPLFYFIMSRLCQD--YDTIGPEG-----GSLKSKLVLQVQATFPENAVTK 1105
Qy 969 WINLTEQWPYRTSWLILYLEETEGIPDQMTLK-----TIYERISK----- 1008
Db 1106 RVKLAQ-----AQVPDELTKLLGNQATFSPIVTVPRRRKHPPIGLR 1151
Qy 1009 -NIPTTKVDVPLEIDGDIRNFVF-----LSSRTPVLVARDKVFLPCTV 1053
Db 1152 IPLPSPWTDNPRDSGEGDTTSLRLCSVIGGTDOAQWEDITGTTKLVYANECANF---TT 1208
Qy 1054 NLDPKREIITADYRAAREQISIGGLAYPPLPHHEGPPRAPSGYSQPPSVCSSTSFNGPEA 1113
Db 1209 NV--SARFWSLDCPRTABAVNFATLLYKEL-----TAVPYMAKFV 1246
Qy 1114 GGVSVPQHSYYSYSGTGPQHPFYNRSGSPAPGPV--VLLNSLNVDAVCE-----K 1162
Db 1247 -----IFAKMNDPREGLRCYCMTDDKVDKDTLQHENFVSEVAR 1284
Qy 1163 LKQIEGLDQSMLEPOYCTTIKANINGRVLAAQCNIDELKEMNMNFGWHILFRSTVLEMRN 1222
Db 1285 SRDIEVLEGMSL-----FAELSG-----NLVPVKAAQOORSFHQSFRENRLAM-- 1328
Qy 1223 AESHVVPEDPRFLSESSSGPAPGEPARRASHNE-----LPHTELS----- 1263
Db 1329 -----PVKVRDSRREBPGSLSFLURKAWKYEDTOHILCHLNITMPPCAKGGADRR 1379
Qy 1264 -SOTPYTLNPFBELNLTGLDEGAPR-----HSNLSWQSQTRTP--SLSSLSQD 1311
Db 1380 RTPTPLALRYSIILSESTPGSLSGTEQAEMKQWAVISEHLGLSMAELARELOFSEVIN--- 1436
Qy 1312 SSIEISKLDKQAEVRYDAYREYIAQMSQL-----EGFGSTTISGRSSPHSTYMGQSSS 1367
Db 1437 -----RIRVENPNSLLEQSVALLNLVWIREG-----QANMENLYTALQSID 1478

QY 1368 GGSIHNLBOEKQKSEPKDDGKSKFLMKRGVDIYSSGCVSTNDASPLDPITBEDE-- 1425
DB 1479 RGEIVNMGSGRQNLKPD-----RRTHDRYSLSPSQMNGYSSL-----QDELL 1525
QY 1426 -----KSDQSGSKL-----LPCKKSSRSSLFQTDLKLKSGSLRYQKLPSDE 1467
DB 1526 SPASLGALSSPURAQYWNEVAVLDAIPLAATEHDTMLEMSQWQVWSAGLTSPSLVTAED 1585
QY 1468 DE---SGTEESDNTPLKDDKORKAGKVERVPKSPHSAEPIRTTFIKAKEYLSDALLDK 1524
DB 1586 SSLECSKAEDSDAT-----GHEWKEGALSEEPRGPE-----LGSLELVEDDVTDS 1631
QY 1525 KDS-----DSGVRSSSESSPHSLNEVADDSQLEKANLIELEDDSHSGKRGPHSL 1576
DB 1632 DATNGLIDLLEQEGORSEKLPGRQDDATGAGQ--DSENEVSLVSGHQGOARITHS-- 1689
QY 1577 SGLQDPFIARMSICSDK-----KSPSECSLIASSPEENW 1611
DB 1690 -----FTVSQVTERSQDLQDWDADGSIYSYLQDAAQGSW 1724

RESULT 7
US-09-949-016-9016
; Sequence 9016, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9016
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9016

Query Match 6.3%; Score 557.5; DB 4; Length 1883;
Best Local Similarity 19.7%; Pred. No. 1.9e-38;
Matches 374; Conservative 286; Mismatches 659; Indels 581; Gaps 72;

QY 21 LKALLEKCKDVRBERCGQTPMLAAEQGNLEIVKELIKNGANCNLEDLNDWTALISASK 80
DB 97 VRELNVYGNVNAQSQKGFPLYMAAQENHLEVVKFLENGANQNVATEDGFTPLAVALQ 156
QY 81 EGHVHIVEELLKGV-----NLEHRDMGCGWTALMWA 111
DB 157 QGHENVVAHLINTGKVKRPLALHIAARNDTRTAVALQNDPNPDVLKSGTFTPLHIA 216
QY 112 CYKGRDVTVELLSHGANSVPT-----GLQY-----SV 139
DB 217 AHYENLVAGLLNRRGASVNTFQNGITPLHIASRRGNVTWRLLLDRGAQIETKQDEL 276
QY 140 YPIIWAAGRHADIVHLLQNGAKVNGSKYGTGTPPLVMAARKHLECVKHLAMGADVD- 198
DB 277 TPLHCAARNHVRISIELLDHGAPIOAKTNGLSPIHMAAQGDHLDVRLLLQYDAEIDD 336
QY 199 -----QEGA-----NSMTALIVAKGGYTSQVKEILKR 226
DB 337 ITLDHLTPLHVAACHGHRVAKVLLDKGAKPNRSGALNGFTPLHACKNHRVWMLLKT 396
QY 227 NPNVNLTKDGNLALMASKEGHEIIVQDILLDAGTYVNIIPDRSGDVTVLICAVRGHVEIV 286

DB 397 GASIDAVTESGLTPLHVASFMGHLPVKNLLQRGASPNVSNVKVETPLHMAARAGTEVA 456
QY 287 RALLQKYADIDIRGQNKNTALYWAKEGNATWVRDILQCNPDTEICTKOGTEPLIKATKM 346
DB 457 KYLLQWKAQVNAKADQDTPHCAARIGHTNMVKLLENNANPLNATTAGTTPHAAARE 516
QY 347 RNIEVELLIDGAKVSAVDKGDTPHIAIRGRSRKLAELLLRNPKDGRLLYRPNKAGE 406
DB 517 GHVETVLALEKEASQACMCKGFTPLHAAKYGKRVAAELLE--RDA-----HPNAAKG 570
QY 407 ---TPYINIDCSHQKSLTQIFGARHLSPTETDGDMLGY-DLYSSALADILSEPTMOPPIC 462
DB 571 NGLTPLHVAHNNLDIVKLLPRGSGSPHAWN--GYTPLHIAAQN-----QVEVA 621
QY 463 VGLYAWGSGKGFLLKKLEDEMKTFAQOIEPLFQFSWLIVFTLLCGG-----LGLLFA 518
DB 622 RSLQYGGSANAESV-----QGVTPH-----LAAQEGHAEMVALLS 659
QY 519 FTVHPNLGIATVSLFLALLVIFFIYVFGRRREGESMNAWLVSTRLARHIGVLELLK- 577
DB 660 KQANGNKGKSLTPLHLV-----AQEG-----HVPVADVLIK 693
QY 578 -----LMFVNPPPELQTTKALPVRELF---TDYNRLSSVG-----GETSLA 616
DB 694 GVMVDATTRMGYTPLHVASHYGNIKLVKFLHQADVNAKTGLGYSPHQAQOQHTDIV 753
QY 617 EMI-----ATLSDACEREFGLA--TRLFRVFKTEDTQSKKKWKTCCPLPSFV 662
DB 754 TLLKNGASPNVSSDGTTPLAIAKRLGYISVTDVLKV--TDET-----SFV 799
QY 663 IFLFIIGCIISGITLLAIFRVPDKHL-----TVNAVLSIASVVG---LAFVLNCRTWQ 714
DB 800 L-----VSDKHRMSPPETVDEIL--DVSEDEGEELSFKARR-----835
QY 715 VLDSLNLNQRKRLHNAASKLHLKSEGFMKVLKCEVELMARMAKTIDSTQNTQRLVVI 774
DB 836 --DSRDVDEKELDFVPKLDQVVESPAIPRIPC-----AMPETVIRSEEQEQASKEY 887
QY 775 DGLDACEQDKVQMLDVTVRVLF--KGPIAIPASDP-----HIIK-----814
DB 888 DEDSLIPSSPATETSDNISPVASPVHTGFLVSPMVDARGSMRSGRHLNGLRVVIPPRTCA 947
QY 815 -----AINONLNSLRDSNINHIDYMN--IVHLVPVFL--NSRGLS 851
DB 948 APTRITCRVLVKOKLSTPPPLAEEEGASRIIAGTGTAQFLSPVIVEIPHAFASHGEG-- 1005
QY 852 NARKFLVTSATNGDVPSCDSTTGIEDADRRVSQNSLGEMTKLGSKTALNRRDITYRRQMQ 911
DB 1006 -DRELVLVLRSENGSV-----WKEHRSRYGSEYLDQILN--GMDEELGSLSELEKKRVC 1055
QY 912 RITITROMSFDLTVKLVLT---EDWFSDISPOTWRRLNIVSVTCRLLRANQISFNWDELAS 968
DB 1056 RIIT--TDFPLXFMVSRILCQD--YDTIGPEG-----GSLKSLKVLVQVATPPENAVTK 1105
QY 969 WINLTQMPVRTSWLLIYLXLEETGIPDOMTLK-----TIVERISK-----1008
DB 1106 RVKLAIQ-----AQPVDELVTKLLGNOATFSPVITVEPRRKRHRPIGLR 1151
QY 1009 -NIPTTKQVEPLEIDGDIRNFEVF-----LSSRTPVLVARDVKVFLPCTV 1053
DB 1152 IPLPSPWNTNPRDSGEGDITSLRLCSVIGGTDAQWEDITGTTKLIVYANECANF---TT 1208
QY 1054 NLDPKUREIIADVRAAREQISIGGLAYPPLPLHEGPPRAPSGVSPQSVCSSTSPNGPFA 1113
DB 1209 NV--SARFMLSDCPTAEAVNFATLLYKEL-----TAVPYMAKFV 1246
QY 1114 GGVSYPQHSSYYSGMTGQHPHPPNRRGSGPAGPV--VLILNSLVNDAVCB-----K 1162
DB 1247 -----IFAKWNPDRGRLCYCMTDDKDKVTKTLEQHENFVEVAR 1284
QY 1163 LKQIEGLDOSMLPOYCTTIKKANINGRVLAQCNIIDELKKEMNNMFGDWHLFRSTVLEMRN 1222
DB 1285 SRDIEVLGMSL-----FAELSG-----NLVPVTKAAQQRSHFQSFRENRLAM-- 1328

Qy	1223	AESHVVPEDPRFLSESSGAPHGEPARRASHNE-----LPHTELS-----	1263
Db	1329	-----PVKVRDSSREPGLSFLRKAMKYEDTQHILCHLNIWTPPCAKGSGAEDRR	1379
Qy	1264	-SQPYTLNPSFBLNTLGLDEGAPR-----HSNLSWQSQTRTP-SLSLNSQD	1311
Db	1380	RTPTFLALRYLSLSESTPGSLSGTEQAEKMAVISEHLGLSLWABELARELOPVEDIN---	1436
Qy	1312	SSIBISKLTDKVAEYRDVREYIAQMSQL-----EGGPGSTTISGRSPHSSTYMGSSS	1367
Db	1437	-----KIRVENPNSLLEQSVALLNLWVIREG-----QNMENMLYALQSID	1478
Qy	1368	GGSHNSLEOKGKDSBPKPDGGRKSLFKMGDVIDYSSSGSVSTNDASPLDPITEEDE--	1425
Db	1479	RGEIVNMLEGSGRGRNLKPD-----RHTDRDYSLSPQMGYSSL-----QDELL	1525
Qy	1426	-----KSDQSGSKL-----LPKKSSESSLSLFTDULKLGSGLRYOKLPSDE	1467
Db	1526	SPASLGALSPPLRADQYMWNEVAVLDALPLAATBHTMLEMSDMQVWSAGLTPSLVTAED	1585
Qy	1468	DE---SGTESDNTPLKDDKORKAEGKVERVPKSPHSABEPIRTFIKAKEYLSDALDK	1524
Db	1586	SSLECSKAEDSDAT-----CHEWKEGALSSEPRGPE-----LGSLELVEDDTVDS	1631
Qy	1525	KDSS-----DSGVRSSSESSPNHSLHNEVADDSQLEKANLIELEDDSHSGKRGIPHSL	1576
Db	1632	DATNGLDLDLEQEBQSGSEELPGSKRQDDATGAGQ--DSENEVSLVSGHQRGQARITHS-	1689
Qy	1577	SGLODPIIARMSICSEDK-----KSPSECSLIASSPEENW	1611
Db	1690	-----PTVSQVTERSQRLQDWDADGSIVSYLQDAAGSW	1724
RESULT 8			
US-09-949-016-9017			
; Sequence 9017, Application US/09949016			
; Patent No. 6812339			
; GENERAL INFORMATION:			
; APPLICANT: VENTER, J. Craig et al.			
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED			
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF			
; FILE REFERENCE: CL001307			
; CURRENT APPLICATION NUMBER: US/09/949,016			
; CURRENT FILING DATE: 2000-04-14			
; PRIOR APPLICATION NUMBER: 60/241,755			
; PRIOR FILING DATE: 2000-10-20			
; PRIOR APPLICATION NUMBER: 60/237,768			
; PRIOR FILING DATE: 2000-10-03			
; PRIOR APPLICATION NUMBER: 60/231,498			
; PRIOR FILING DATE: 2000-09-08			
; NUMBER OF SEQ ID NOS: 207012			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 9017			
; LENGTH: 1883			
; TYPE: PRT			
; ORGANISM: Human			
US-09-949-016-9017			
Query Match 6.3%; Score 557.5; DB 4; Length 1883;			
Best Local Similarity 19.7%; Pred. No. 1.9e-38;			
Matches 374; Conservative 286; Mismatches 659; Indels 581; Gaps 72;			
Qy	21	LKALLECKDVERNECGTFLMAAEOGNLEIVKELIKNGANCNLEDLNNWLTALISASK	80
Db	97	VRELNVNGANVAQSQKGFTFLYMAAQENHLEVVKFLLENGANCNATEDGFTPLVALQ	156
Qy	81	EGHVHIVEELLKCGV-----NLEHRDMGCGWTALMWA	111
Db	157	QGHENVVAHLINYGTKVKVRLPALHIAARNDDRTAAVLLQNDPNPVLKGTFTPLHIA	216
Qy	112	CYKGRDVTWVLLSHGANPSVT-----GLQY-----SV	139

Db	217	AHYENLVNAQQLLNRGASVNFPTQNGITPLHIASSRRGNVIMVRLLLDRGAQIETKDEL	276
Qy	140	YPIIWAAGRHADIVHLLQNGAKVNCSDKYGTTPLVWAARKGHLECKVHLLAMGADV-	198
Db	277	TPLHCAARNGHVRISILLDHGAPIQAKTKNGLSPIHMAAQGDHLDCCVRLQLQDAE	336
Qy	199	-----QEGA-----NSMTALIVAVKGGYQTSVKELIKR	226
Db	337	ITLDHLTPLVAAHCHHRAKVKLLDKGAKPNSRALNGFTPLHTACKKNHVRVMEILLKT	396
Qy	227	NPNNVLTDKDGNATMIASKEGTEIVODLDAQTYNIPDRSGDTVLIGVRGHHVIV	286
Db	397	GASIDAVTESGLTFLHVASFMGHLPIVKNLLQRGASPNVSVKVTETPLHMAARAGHTEVA	456
Qy	287	RALLOKVDADIDIRQDNKNTALYNAVEKGNATWVDILQCPDTEICTKOGETPLIKATKM	346
Db	457	KYLLQNKAKVNAKADQDTPLHCAARIGHTNMVKLLLENNANPNLATTAGHTPLHIAARE	516
Qy	347	RNIEVVELLDKGAKVSAVDKGDTPHIAIRGSRKLAELLLRNPKDGRLLYRPNKAGE	406
Db	517	GHVETVLALLEKEASQACMTKKGFTPLHVAAKYKVRVAELLBS--RDA---HPNAAGK	570
Qy	407	---TPYNDCHSHOKSIUQTIFGARHLSPTETDGMILGY-DLYSSALADILSEPTWQPPI	462
Db	571	NGLTPLHVAVHNHNLDIVKLLPRGSGPSHPAWN--GYTPLHIAAKQN-----QVEVA	621
Qy	463	VGLYAQWGSCKSLFKLEDEMKTFFAGQIETPLFPQPSWLVFLTLCLCGG---LGLLFA	518
Db	622	RSLLQYGGSANASV-----QGVTPHL-----LAAQEGHAEMVALLLS	659
Qy	519	FTVHPNLGIAVSLSFALLIYFFIVYFGRRGESNMAWLVSLRLARHIGYLELLIK-	577
Db	660	KQANGNLGNKSLTPLHLV-----AQEG-----HVPVADVLKX	693
Qy	578	-----LMFVNPPPEQTTKALPVRFLF---TDYNRLSSVG-----GETSLA	616
Db	694	GVNVDAATTRMGYTPLHVASHYGNIKLVKFLHQADVNAKTKLGYSPLHQAQGHDTIV	753
Qy	617	EMI-----ATLSDACEREGFLA-TFLFRVFKTEDQGGKKWKKTCLPSFV	662
Db	754	TLLLKNGASPNVSSDGTTPLAIAKRLGYISVTDVLKV-TDET-----SFV	799
Qy	663	IFLFIIGCIISGITLALFAIVDPKHL-----TNAVLISIASVVG---LAPVLNCRWTWQ	714
Db	800	L-----VSDGRMSFFETVDEIL-DVSEDEGEELISFKAER-----	835
Qy	715	VLDLSLNSQRKRLHNAASKLHKLKSEGFMKVLCVELMARMAKTIDSTFQNTQRLVVII	774
Db	836	--DSRDVDEEKELLDVFKLDQVVESPAIRIPC-----AMPETVIVIRSEEQOASKEY	887
Qy	775	DGLDACEQDKVLQMLDTRVLFPS--KGPFTAIASDP-----HIITK-----	814
Db	888	DEDSLIPSSPATETSDNISPVASPVHTGFLVSVFMDARGSGMRSGRNLGRVLPPTCA	947
Qy	815	-----AINQNLNSVLRDSNINGHDYMRN-IVHLPVFL-NSRGLS	851
Db	948	APTRITCLVVKPKLSTPFPPLAEEGLASRIIALGPTGAQFLSPVIVIEIHFASHGRG--	1005
Qy	852	NARKFLVTSATNGDVPSCDITGIEDADRRVSNQSLGEMTKLGSKTALNRDRTVRRQM	911
Db	1006	-DRELVLRLSENGSV-----WKEHRSRYGESYLDQILN-GMDEELGSLLEEKKRVC	1055
Qy	912	RTITRQMSPLTKLAVT---EDWFSDISQPMRLLNIVSVTGELLRANQISFNWDRLAS	968
Db	1056	RIIT--TDFPLYFVIMSRLCQD-YDTIGPEG-----GSLKSKLVLVQATFPENAVTK	1105
Qy	969	WINLTEQWPYRTSWLILYLEETEGIPQDMLK-----TIVERISK-----	1008
Db	1106	RVKLAQ-----AQPVDELTKLLGNQATFSPITVTPRRKFRHPIGLR	1151
Qy	1009	-NIPTTKQVPELLEIDGDIRNFVF-----LSSRTPVLVARDVKVFLPCTV	1053
Db	1152	IPLPSPWTDNPRDSGEDTTSLRLCSVIGGTDAQOWEDITGTTKLVYANECANF---TT	1208

QY 1054 NLDPKLEIADYRAAREQISIGLAYPPLPLHEGPRAPSGYSQPPSCSSTSFNGPFA 1113
Db 1209 NV--SARFWSLDCPRTAEAVFATLLYKEL-----TAVPYMAKFV 1246
QY 1114 GGVSPQPHSSYYSGMTGPHFPYNGSGPAPGV--VLINSLNVDACE-----K 1162
Db 1247 -----IPAKNDPREGRACYCMTDDKVKTYLEOHENFVEVAR 1284
QY 1163 LKQIEGLDQMLPOYCTTIKANINGRVLAOCNIDELKEMNNFGDWHLFRSTVLEMRN 1222
Db 1285 SRDIEVLEGMSL-----FABELSG-----NLAVPKAAQORSFHFQSFRENRLAM-- 1328
QY 1223 AESHVVPEDPRFLESSESSGAPGEGEPARRASHNE-----LPHTELS----- 1263
Db 1329 -----PVKRDSSREFGGSLFLRKAMKYEDTQHLCHLNLTMPPCAKGSGAEDRR 1379
QY 1264 -SOTPYTLNFSPEELNLTGLDECAPR-----HSLNWSQSTRTTP-SLSSLSNQD 1311
Db 1380 RPTPLALRYSLSESTPGSLSTEGEAEMKMAVISHLGLSWAELARELOFSVEDIN--- 1436
QY 1312 SSIEISKLDKQVARYDAYREYIAQMSQL-----EGPGSTTISGRSSPHSTYMGQSSS 1367
Db 1437 -----RIRVENPNSLLEQSVALLNLWVIREG-----QANMENLYTALQSID 1478
QY 1368 GCSIHNSIQEKGKDEPEDDGKSFMLKRGVDIYSSGVSNTDASPLDPITEDE-- 1425
Db 1479 RGEIVNMGSGRQSNLKP-----RRHTDRYSLSPSQMNGYSSL-----QDELL 1525
QY 1426 -----KSPQSGSKL-----LPGKKSERSSLFQTDLKLKGLGRYQKLPDSB 1467
Db 1526 SPASLGCALSSPLRADOYNEVAVLDAIPLATEHTMTLEMSQWVSAGLTPSLVTAED 1585
QY 1468 DE---SGTESDNTPLKDKKRAEGKVERVPKSPHSAEPIRTPIKAKEYLSDALDK 1524
Db 1586 SSLECSKAEDSDAT-----CHEWKLEALSEEPRGP-----LGSLELVEDDTVDS 1631
QY 1525 KQSS-----DSGVRSESSPHSLNEVADDSOLEKANLIEEDDSHGKRGIPHSL 1576
Db 1632 DATNGLIDLLEBEGQRSEKLPGRQDDATGAGQ-DSENEVSLVSGHQRGQARITHS- 1689
QY 1577 SGLQDPPIIARMSICSBDK-----KSPSECSLIASSPEENW 1611
Db 1690 -----PTVSQVTERSQDRQLQDWADGSIYSYLQDAAQGSW 1724

RESULT 9

US-09-949-016-6964
; Sequence 6964, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6964
; LENGTH: 1856
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6964

Query Match 6.3%; Score 555.5; DB 4; Length 1856;
Best Local Similarity 19.7%; Pred. No. 2.7e-38;

Matches 374; Conservative 286; Mismatches 659; Indels 581; Gaps 72;
QY 21 LKALLEKCKVDNERNECGOTPLMIAAEOGNLEIVKELIKNGANCNLEDNDWTALLSASK 80
Db 94 VRELNYGANVAQSGKGTPLYMAAENHLEVYKFLLENGANQNATEDGFTPLVALQ 153
QY 81 EGHVHIVBELLLKGV-----NLEHRDMGWTALMWA 111
Db 154 QGHENVAHLINVTGKVRLLPALHIAARNDTRTAARVLLQNDPNDVLSTGFTPLHIA 213
QY 112 CYKGRTDVVVELLSHGANSVT-----GLQY-----SV 139
Db 214 AHVENLVAQLLNRGASVNFTPQNGITPLHIASRRGNVIMVRLLLDRGAIETKTDEL 273
QY 140 YPIIWAAGHGHADIVHLLQNGAKVNCSDKYGTTPLVWAARKGHLSCVXHLWANGADVD- 198
Db 274 TPLHCAARNGHVRISILLDHGAPIQAKTKNGLSPIHMAAQGDHLDVLRLLQYDAEIDD 333
QY 199 -----QEGA-----NSMTALIVAKGVYQTSVKELKR 226
Db 334 ITLDHLTPLHVAHCHHRVAKVLLDKGAKPNRSALNGFTPLHIACKKHVRVWELLKT 393
QY 227 NPNVNLTKDGNATMIASKEGHEITVQDLLDAGTYVNIIPDRSGDTVLICAVRGHVEIV 286
Db 394 GASIDAVTESGLTPLHVASFMGHLPVKNLLQRGASPNVSNVKTPLHMAARAGHTEVA 453
QY 287 RALLQKYADIDIRQDNKTALYWAKEGNATWVRDILQCNPDTEICTOGETPLIKATKM 346
Db 454 KYLLQKAKVNAKADQDTPLHCAARIGHTNMVKLLLENNANPNLATATTAGHTPLHAAE 513
QY 347 RNIEVVELLDGAKVASVDKGDTPHHTAIGRSRSEKLAELLRNPKDGLLRVPHKAGE 406
Db 514 GHVETVLALEKEASQACMTKKGFTPLHVAAKYKVRVAELLE--RDA-----HPNAGK 567
QY 407 ---TPYINDCSHQKSILTQIFGARHLSPTETDGMGLGY-DLYSSALADILSEPTMOPPIC 462
Db 568 NGITPLHVAVHNNLDIVKLLPRGSGSPHS PAWN--GYTPLHIAKQN-----QVEVA 618
QY 463 VGLYAWQSGKGFLLKLEDEMKTFAQOIEPLFPQSWLIVFLTLCCG-----LGILFA 518
Db 619 RSLQYGGSANAEV-----QGVTPH-----LAAQEGHAEMVALLS 656
QY 519 FTVHPNIGIAVSLFLALVIFPIVIYFGRREGESMNWAVLSTRLAHIGVLELLK- 577
Db 657 KQANGNLGNKSGLTPHLV-----APEG-----HVPVADVLIK 690
QY 578 -----LMFVNPPPELPTQTKALPVRELF---TDYNRLSSVG-----GETSLA 616
Db 691 GWMVDATTTRMGYTPLHVASHYGNIKLVKELLQHQADVNAKTKLGYSPHQAAQGHDTIV 750
QY 617 EMI-----ATLSDACEREFGLA-TRLFRVFKTEDTQKKKWKTCCLPSPFV 662
Db 751 TLLKNGASPNVSSDGTTPLAIAKRLGYISVTDVLKV--TDET-----SFV 796
QY 663 IFLFIIGCIISGILLAIERVPKHL-----TVNAVLSIASVVG---LAFVLCNRTWQ 714
Db 797 L-----VSDKHMSFPETVDEIL-DVSEDEGEELSFKAE-- 832
QY 715 VLDSLNSQRKRLHNAASKHLKSEGFMKVLKCEVELMARMAKTIDSTQNTRLVWII 774
Db 833 --DSRDVDEKELLDFVPKLDQVVESPAIPRIPC-----AMPETVIRSEEQEQASKEY 884
QY 775 DGLDACEQKVLOMLDTRVILFS--KGPFIAPASDP-----HIILK----- 814
Db 885 DEDSLIPSSPATETSDNISFVASPVHTGFLVFMVDARGSMRSGSRHNGRLRVVIPPRTCA 944
QY 815 -----AINONLNSLRDSNINGHYMRN-IVHLPVFL-NSRGLS 851
Db 945 APTITCRLVKQKSLTPPPLEBEEGLASRIIALGTGTAQFLSPVIVEIPHFAHSG-- 1002
QY 852 NARKFLVTSATNGDVPSCDITGIEDADRRVRSQNSLGMTKLSKALTANRRDTRRROMQ 911
Db 1003 -DRELVLRSNGSV-----WKEHRSRYGESYLDQILN-GMDEELGSLEBKRCVC 1052

Qy	617	EMI-----ATLSDACERERGFIA-TRLFRVFKTEDTQGGKKWKTKCCLPsfv	662
Db	751	TLKNGASPNVSSDGTTLAIKRLGYISVTDVLKw-TDET-----SFV	796
Qy	663	IFLFIIGCIISGITLLAIFVDPKHL-----TVNAVLIStASVVG---LAFVLNCRtWQ	714
Db	797	L-----VSDKHRMSPFETVDEIL-DVSEDEGBELISFKAEr	832
Qy	715	VLDSLNSQRKLHNAASKLUHKSEGMVKLVKCEVELMARMAKTIDSFTQNTRLVvII	774
Db	833	--DSDRDEEKELLDFVKLDQVVESPAIPrPC-----AMPETVVIRSEEGEQASKEY	884
Qy	775	DGLDACEQDKVLQMLDTRVVLFS--KGPFTAIrASDP-----HIIK-----	814
Db	885	DEDSLIPSPATETSDNISPVASPVHTGFLVFMVDARGGSMGRSHRGLRVVIPPtCA	944
Qy	815	-----AINONLNSVLrDSNINGHDYMRN-IVHLrPVFL-NSRGLS	851
Db	945	APrTITCrLVKPKLSTPPPLABEEGLASRIIALGPTGAQFLSPVIEIHPFASHGrG--	1002
Qy	852	NAKFLVTSATNGDVPSCDGTGTIQEDADRRVQNSLGEMTKLGSKTALNRDRDYRRRQMQ	911
Db	1003	-DRELVLVRSENGSV-----WKEHRSYGESYLDQILN-GWDElGSLEELKKrVC	1052
Qy	912	RTITRQMSFDLTkLLVrT---EDWFSDISPQrMRLNIVSVTgELLERANQISFNWDELAS	968
Db	1053	RIIT--TDFFLYFVIMSRlCOD-YDIIGrEG-----GSLKGLKVLVQATrPENAVTK	1102
Qy	969	WINLrTEQWPVrTSLWILYLrETEgIPQOMTLK-----TIVrERISK-----	1008
Db	1103	RVKLALQ-----AQrPVDELVTKLlGNOATrSPIVrVEPrRRKFrHPrIGLr	1148
Qy	1009	-NIPTTKDVBLElDIGDINrEVrF-----LSSrTPVLVARDVKVlFPCrV	1053
Db	1149	lPLPpSTWDNPrDSGEgDTTLRLlCSVIGGTQAOwMEDITGTTLKYANECANrF--TT	1205
Qy	1054	NLDPLREIADVRAAREQISIGGLAYPPLPLHEGPrPRArSGVrSQPpSVCSrTFNGCPFA	1113
Db	1206	NV--SARFVUSDcPrTAEANrFATLLYKEL-----TAVPYMAKFrV	1243
Qy	1114	GGVVSrPOPpSHSYSGMTGPQHPFNrGRSGPAGrPV--VLLNSLNVDAVCE-----K	1162
Db	1244	-----IFAKMNDPrEGrLRCYCMrTDDKVDKTLQHENrFVEVAr	1281
Qy	1163	LKOIEGLDQMSLrPOYCTTIKANINGrVLAQCNlDELKEMNNMFGDWHLrFRSTVLEMRN	1222
Db	1282	SRDIEVLGMSL-----FABLrSG-----NLrPVVKAAQORSHrFQSFRENrLAM--	1325
Qy	1223	AESHVrPEDPrFLrSESSSGPAPHCEPrARRASHNE-----LPHrTELS-----	1263
Db	1326	-----PVKVRDSrREPGGSLSrFLRKANKYEDrTOHILCHLNIrTTPPCAKSGGAEDRR	1376
Qy	1264	-SOTPYTLNFrSFBELNTlGLDEGAPr-----HSNLrSWOSQrTRrTP-SLSSLSNQD	1311
Db	1377	RTPTPLALrYSIIrSESTPGSLrSGTQEAKrMNAVISEHLGLSWEALARElQrFVEDIN--	1433
Qy	1312	SSrEISKlTDKVAErDAYrDArVrIAQMSQL-----EGGPGGTTrISGrSSPHSTrTYMGSSS	1367
Db	1434	-----RIRVENrPNSLrEOSVALLNLrWIrREG-----QNaNMENrLYALQrSID	1475
Qy	1368	GGSlHNSLEQKGKDSrPKDDGrKSLFKMRGDVIDYSSGSVSTNDASPLDPrITEEDE--	1425
Db	1476	RGEIVNlMGSGrQSRNLKPD-----RRHTDrDYSLrSPQMWGYSSL-----QDELL	1522
Qy	1426	-----KSDQSGSL-----LPCKGSrSSSLrFQTLKLGSGSLrYQKLPSDE	1467
Db	1523	SPASlGCALSSPrADQYrWNEVAlDAIPLAArTEHDTrMLESMDQVWSAGLrTSPSLVrTAE	1582
Qy	1468	DE---SGrTESDNTPLKODrKDKrAEKGrVrVPKSPHrSEArPTFrTFKAEYLSDALLDK	1524
Db	1583	SSlECSKAEDSDAT-----GHEWKLEGALEEPrGPGE-----LGSLElVEDDTVD	1628

```

Qy 1525 KDS- - - - - DSGVRSSESSPHSLHNEVADDSOLEKANLIELEDDSHGKRCIPHSL 1576
      :: : ||| : | : : : : : : : : : : : : : : : : : : : : : : :
Db 1629 DATNGLLDLLEBEGQRSEEKLPGSKRKDDATGAGQ-DSENEVLVS GHQGRARITHS- 1686
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1577 SGLQDPPIIARMSICSEDK- - - - - KSPSECSLIASPBEENW 1611
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1687 - - - - - PTVSQVTESQDELQDWDADGSIVSVYLQDAAGQS 1721
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
US-09-949-016-7659
; Sequence 7659, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7659
; LENGTH: 2753
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7659

```

Query Match 6.28; Score 549; DB 4; Length 2753;

```
QY 553 ESNWAWVLSTRLARHIGYLELL--LKLM----- 579
Db 662 TTANGNTALA--IAKGLYISVVDTLKVVEEVTITTEKHKLNPETMTVELDVSD 719
QY 580 -----FVNPPPEQTTKALP-VREL-FTDYNRLSSVGGETSLAEMIATLSD 624
Db 720 EGDGDTMTGGEYLRPEDKELGDDSLSPSQFLDGMVYRLSLEGRSDSDRSHTLSH 779
QY 625 AC-----ERFEGFATRLFRVFKTEDTQKKK-- 860
Db 780 ASYLDRSAVMDSDVWIPSHQVSTLAK-----EAERNVRLSMGTENLNVALLSSPIHS 833
QY 661 FVIFLFI-----GCIISGITLAI FRVDPKHLTVNAVILISIASVGLAFVLCRTW 712
Db 834 GFLVSPMVDARGAMRGCRHGLRII-----IPRCKTAPT-----RVTCL- 875
QY 713 WQVLDLLNSQRKL-----HNAASKLHLKSEG-----FMKVLKCEV 750
Db 876 -----VKHRLATMPMVEGEGLASRLIEVGSQAQFLGPVIVEIPHFAALRGKER 926
QY 751 ELMARMAKTIDSTQN-----QTRLVVIIDGLDACEQDKVLOMLDTRVFLFSKGPFIATP 805
Db 927 ELVLRSENGDSWKEHCFDYTEDELNEILNGMD-----EVLDSGEDLEKKRICRIIT 978
QY 806 ASDPH--IIIKAINQNLNVLSDNSNINGHDYMRNIVHLFPVFLNSRGLSNARKFLVTSATN 863
Db 979 RDPQPFVAVSRIKQ-----DSNLIGPE-----GGVLSSTVVPQVAVFPPEGALT 1023
QY 864 GUVPCSDTTGIOEDADRRVSQNSLGBMTK--LGSKTALNRDFTY--RRROMORTIFROMS 919
Db 1024 KRI-----RVGLQ-----AQMHSSELVKILGNKATFPIVLEPRRRKFKPIWTWTP 1072
QY 920 FOLTCLLVTDWFSDISPOTMRRLNIVSVTGRLRANQISFNWDRLASWINLT---EQW 976
Db 1073 VPKASDVMNGFGGDAP-TLRL--CSITGTTTPA-----QWEDITGTTPLTFVNECV 1123
QY 977 PYRTS-----WLILYLEETEGIPDQMTLXIYERISKNIPYTKDVPLEIDGDIRNFV 1031
Db 1124 SFTTNVSAFRLWI-----DCRQIQESVTFASQVYREIICVPY-----MAKFV 1166
QY 1032 FLSSRTPVLVARDVKVFLPCTVNLDPKL-----REIADVRAARQISIGGLAYPLPL 1085
Db 1167 FAKSHDP-----IEARLCFCMTDDKVKLTLEQOENFAVARSRD-----VEV 1209
QY 1086 HEGPPRAPSGYGQPPVSCVSTPNGP-----FAGGVVSPQHSYYSYSGMTGPQHPF-- 1136
Db 1210 LEGKPIYVDFGNLVLTKSTRYDSGLTRTTIIEGAKIHERPGASSYLQ---KHFAD 1266
QY 1137 -YNRSGPAPGPV---LLNSLVNDVACEKLQIEGLDQSMPLPOYCTTIKKNINGRVL 1191
Db 1267 LYKDDTESETSVLKSHLVNEVPVLASPOLISEVSEMQDLTKM--TAILTVDSDKA- 1323
QY 1192 AQCNIDELKEMMNFQDWHLFRSTVLEMRNABSHVVVPDPRFLSB-SSSGPAPHCEPAR 1250
Db 1324 GSIKVKELKAAEEBGE-----PFEIVER-----VKDEKVNILRSIGCTRDRESS 1372
QY 1251 RASHNELPHTLESSQPTTYTLNFSFEELNTGLDEGAPRHSNLSWQSQTRTPPSLSSLSNQ 1310
Db 1373 QSSRSERGLVE--EEMVIVSDEIEEAR-----QKAPLEITEYPCVEVIDKEIKGVKEK 1425
QY 1311 DSSIEISKLTDKQVARYDAYREYIAQMSQLEGGPGSTTISGRSSPSTHYMGQSSSGGS 1370
Db 1426 DSTGLNYLITDNLNT-CVPLPKELQTVQDKAGKCEALAVGRSS----- 1469
QY 1371 IHSNLQEKKGQSEP-----KPDGRKSFMLKRGDVIDYSSSGVSTNDASPLDPI 1420
Db 1470 -----EKEGKILPPDETOSTQKHPSIGIKKPVRRK-----LKEK 1505
QY 1421 TEEDEKSDQSKLLPGKKSRSRSLFQDQLKLGSGLYRQKLPSEDESSTGTESDNTPL 1480
Db 1506 QKQKEGLQASAEKAEKKGSSBESLGE-----DPLAPEPLPTVKATS-----PL 1551
```

```
QY 1481 LKDDKDKRKAEGKVERVPKSP-----HSAEPIRTFIKAEVYLSDALLDKKDSSDSGVRSS 1535
Db 1552 IBETPIGSIKDKVKALQKRVEDEQGRSKLPIR--VKGKEDVPKTKTHRPHPAASPSLKS 1609
QY 1536 E-----SSPNHLSHNEVADDSQLEKANILELDDSHSGKRGIPHSLSGLQDPIIARMSI 1589
Db 1610 ERHAPGSPKPTERHSTLSSSAKTERHPVVS-----PSSKTEKHSVP---SPS 1654
QY 1590 CSBEDKKSPECSLIASSPENWPAQKAYNLNRTPTVTILNNSAPANRANQNDFDEMEGI 1649
Db 1655 AKTERHSPASS--SKTEKHSVP-----SPSTKTERHSPVSTTERH---PPV 1698
QY 1650 RTSOVILRPSSSPNPTTIONENLK-SMTHKRSQSSSYTRLSKDDP 1694
Db 1699 SPSGKTKDRKPPVSPSGRTEKHPVSPGRTEKRLPVPSPGRTDKHQP 1744

RESULT 13
US-09-949-016-7660
; Sequence 7660, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7660
; LENGTH: 2753
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7660
```

Query Match 6.2%; Score 549; DB 4; Length 2753;
Best Local Similarity 21.2%; Pred. No. 2.2e-37;
Matches 392; Conservative 284; Mismatches 710; Indels 460; Gaps 72;

```
QY 21 LKALKEKCDVDERNECGQTPLMIAAOGNLEIVKELIKNGANCNLEDLDNWTALISAK 80
Db 187 VELLERGAFLARTKNGLSPLHMAAQGDHVECVKHLQHKAPVDDVTLDTLTLHVAAH 246
QY 81 EGHVHIVELLKCGVNLHRDNGGWTALMWACVKGRTDVVELLSHGAN-PSVTGLQYSV 139
Db 247 CGHYRVTKLLDKRANPNARALNGFTPLHACKNRKIVMELLYKVGASIQAIT--ESGL 304
QY 140 YPIIWAAGRHADIHVHLLQNGAKVNCSDKYGTTPLVAAARKGHLECVKHLAMGADVQ 199
Db 305 TPIHVAAPFMGLNIVLLLLQNGASPDVTNIRGTALHMAARAGQVEVRLLENGALVDA 364
QY 200 EGANSMTALIVAKGYTQSVKEILKRNPNVNLTKDGNLTALMIASKEGTEIVQDLDA 259
Db 365 RAREEQTPHLIASRLGKTEIVQLLLQHMAHPDAAATNGYTPHLHSAREQVDVASVLEA 424
QY 260 GYVNI PDRSGDTVLIGAVRGHVEIVRALLQYADIDIRGQDNKTALTVAVEKGNATWV 319
Db 425 GAAHSLATKKGFTPLHVAKYGLDVAKLLQRRRAAASAGKNGLTPLHVAHYDNQKVA 484
QY 320 RILQCNPTTEICTKGETPLIKATKMRNIEVVELLDKGAKVSADVKKGDTPLHIAIRG 379
Db 485 LLLLEKGAIPHATANKNYTPHLIAKQNMQASTLLNTGAETNIVTKQGTPLHLASOE 544
QY 380 RSRKLAELLLRNPKQGRLLYRPNKAGETPNYIDCSHQK----SILTOIFGARHLSPTTD 435
Db 545 GHTDMVTLLL---DKGANIHMSTKSLTSLHLLAAQEDKVNADVILTK-HGADQDAHTK-- 598
```

```
Qy 436 GDMGLGYDLYSALADIIILSEPTMQPPIICVGLYAOWGSGK--SFLLLK-LEDEMKTFAGQOI 492
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
599 ---LGY-----TPLIVACH--YGNVKNVFNLLKQGANVNAKTNG--Y 634
Qy 493 EPLQFQSWLIVFLVLLCGGLGLLPAFTVPHNLGIAVLSLFLALYIPFVIYFGRRGREG 552
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
635 TPLHQ-----AAQGGH-----THIINVILLOHGAKPNA 661
Qy 553 ESNWMAWLSRLARHIGYLELL--LKLK-----579
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
662 TTAGNTALA--IAKRLGISVVDTLKVVTBEVTTTTTTEKHKLNVPTMTVEVLVDSD 719
Qy 580 -----FVNPPPELPEQTTKALP-VRFPL-FTDYNRLSSVGGGETSLAEMIATLSD 624
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
720 EEGDDTMTGDGGEYLRPEDLKELGDDSLPSSQFLDGMNLYRLSYLEGGSSDSSDRSHLISH 779
Qy 625 AC-----EREGFLATLRFVRKFTEDTQKKKW-----KKTCCPLPS 660
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
780 ASYLURDSAVMDDSVVIPSQVSTLAK-----EARNSTYRLSMGTENLDNALSSSPIHS 833
Qy 661 FVIFLPII-----GCILSGITLLAIFRVDPKHLTVNAVLISIASVVGFLAVLNCRTW 712
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
834 GFLVSFVWDARGGMRCGRHNGRLII-----IPPRKCTAPT-----RVTCLL- 875
Qy 713 WQVLDSLINSQRKL-----HNAASKLHLKLESG-----FMKVLKCEV 750
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
876 -----VKRHLATMPWPVEGEGLASRLIEVGPSCAQPLGPVIVEIPHFAALRGKER 926
Qy 751 ELMARMAKTIDSTFQN-----QTRLVVIIDGLDACEQKVQLQMLDVTVRVLPFSKGPFFIAIF 805
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
927 ELVVLRSENGDSWKEHFCYDEDELNEITLNGMD-----EVLDSPEDELEKKRICRIIT 978
Qy 806 ASDPH--LIIKAINONLSNVLDSNINCHDYMRNIVHLVPVPLNSRGLSNARKFLVTSATN 863
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
979 RDPQYFAVSRIRKQ-----DSNLIGPE-----GGVLSSTVVPQVQAVPEGALT 1023
Qy 864 GDVPCSDTTGIEDADRRVONSIGEMTK--LGSKTALNRDVTY--RRRQWORTITROMS 919
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1024 KRI-----RVGLQ-----AQPMHSELVKILGNKATFSPVITLEPRRRKFKHKEITWIP 1072
Qy 920 FDLTKLIVTEDFWSDISQPMRRLNLNIVSVTGRLLRANQISFNWDRLASMINLT---EQW 976
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1073 VPRASSDVMNLGFGGDAP-TLRLI---CSITGGTTPA-----QWEDITGTTPLTFFVNECV 1123
Qy 977 PYRTS-----WLLIYLETETGIPQOMTLKTIYERISKNIPPTKDVPELLEIDGDIRNFV 1031
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1124 SFTTNVSARFWLI-----DCRQIQESVTFASQVYREIICVPY-----MAKFFV 1166
Qy 1032 FLSSRTPLVARDVKVFLPCTVNLDPKL-----REIIADVRAAREQISIGLAYPPLPL 1085
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1167 FAKSHDP-----IEARLCFCMTDDKDKTLEQOENFAEVARSD-----VEV 1209
Qy 1086 HEGPPRAPGSQSPVCSSTSFNGP-----FAGGVVSPQPHSSYSYSGMTGQHPFP-- 1136
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1210 LEGKPIYVDCFGNLVPLTKSTRYDSGLTRITIIIEGAKIHERPCASSYVQL---KHFPAD 1266
Qy 1137 -YNRGSPAPCPVV-----LLNSLNVDAVCEKAKQIEGLDQSMLPQYCTTIKANANRVL 1191
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1267 LYKXDETESTETSVLKSHLVNEVPLASPDLLSEVSEMKQDLIKM--TAILTTDVSOKA- 1323
Qy 1192 AQCNIDELKEMNMNFGDWHLFRSTVLEMRNAESHVVPEDPRFLSE--SSSGPAPHGEPAR 1250
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1324 GSIVKELVAAEEPEGE---PPEIVER-----VKEDLEKNEILRSGTCRTDESSV 1372
Qy 1251 RASHNELPHTELSSQTPYTLNFSPEELNTLGLDEGAPRHSNLSQSQTRTPSLSSLSNSQ 1310
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1373 QSSRSRGLVE--BEWVIVSDEEIEEAR-----QKAPLEITEYPCVEVRIDKEIKGVKEK 1425
Qy 1311 DSSIEISKLTDKVOAERYDAYREVIAQMSQLEGPGSGTITISGRSSPHSTYMGQSSSGGS 1370
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1426 DSTGLVNYLTDLLNT-CVPLPKQLOTVQDKAGKCKEALAVGRSS-----1469
```

```
Qy 1371 IHSNLEQEKGDSEPP-----KPDGRKRSFLMKRGDVIDYSSSGVSTNDASPLDPI 1420
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1470 -----EKEGKOIFPDETQSTQKHKPSGLGKPKVRRK-----LKEK 1505
Qy 1421 TEDEKSDQSGSKLLPKKKSRSRSLFQTDULKGSGRLRYOKLPSDEDESGTEESDNTPL 1480
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1506 QKQKEEGLQASAEAKELKGSSEESLGE-----DPGLAPEPLPTVKATS-----PL 1551
Qy 1481 LKODKDRKAEGKVERVPKSP-----HSAEPITRFIKAEVLSLALLDKDSSDSGVRS 1535
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1552 IETPIGSIKKVAKALQKRVDEQKGRSKLPIR--VKGEDVPKTTTHRPHAPASPSLK 1609
Qy 1536 E-----SSPNHSLHNEVADDSQLEKANLIELEDDSHSGKRGIPHSLSGLQDPIIARMSI 1589
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1610 ERHAPGSPSPKTERHSTLSSSAKTERPPVS-----PSSKTEKHSVP---SPS 1654
Qy 1590 CSEDKKSPSCSLIASSPEENWPAQKAYNLNRPSTVTLNNSAPANRANQNFDMEGI 1649
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1655 AKTERHSPASS--SKTEKHSVP-----SPSTKTERHSPVSVSTKTERH---PPV 1698
Qy 1650 RETSQVILRPSSSPNPTTIONENLK-SMTHKRSQSSVYRLSKDPP 1694
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
1699 SPSGTDKRPVPSGSGTEKHPPVSPORTEKRLPVLPSGRTDKHP 1744
```

RESULT 14

```
US-09-538-092-1246
; Sequence 1246, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurapatSeqformatter Version 0.9
; SEQ ID NO 1246
; LENGTH: 3924
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q01484
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q01485
; US-09-538-092-1246
```

```
Query Match 6.1%; Score 539; DB 4; Length 3924;
Best Local Similarity 20.1%; Pred. No. 3.3e-36;
Matches 421; Conservative 294; Mismatches 717; Indels 662; Gaps 81;

Qy 14 EENIPALKALLEKCKDVDERNECGQTPLMIAAQGNLEIVKELIKNGANCNLEDLDMWT 73
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
73-KEGHVGLVQELLGRGSSVDSATKKGNTALHIALAGQAEVVKVLVKEGANINAOQNGFT 132
Qy 74 ALISASKEGHVHIVEELLKCGVN-----LEHRDMG---103
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
133 PLMAAQENHIDVVKYLLENGANGQSTATEDGFTPLAVALQOQHNOQVAAILLENDTKGVR 192
Qy 104 -----GWTALMWACYKGRDVTVELL 123
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
193 LPALHIAARKDDTKSAALLLQNDHNAVQSKMMVNRTEGFTPLHIAAHYGVNVNATLL 252
Qy 124 LSHGANPSVTGLQYSVVPIIWAAGRGHADIVHLLLONGAKVNCSDKYGTTPLVVAARKG- 182
Db      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

```

Db 253 LNRGAADVFTA-RNGITPLHVASKRGNTNNVKKLLDRGGQIDAKTRDGLTPLHCAARSGH 311
Qy 183 -----HLECVKHLAMGADVDOEGANSMTALIV 210
Db 312 DOVVELLLERGAPELLARTKNGLSPLHMAAQGDHVECVKHLQKAPVDDVTLTYLTALHV 371
Qy 211 AVKGGYTQSKEITLKENPNVNLTKDGNLTALMASKE----- 247
Db 372 AAHCGRYRVTKLLDKRAHPNARALMGFTPLHIAACKNRKIKWELLVYKYGASIQAITESG 431
Qy 248 -----GHTEIVQDLIDAGTYNNIPDRSGDTVLIGAVRGHVEIVRALIQKYADID 297
Db 432 LTPHVAAFMGHLNIVLLLLQNGASPDVTNIRSETALHMAARAGQVYVRCLLRNGALVD 491
Qy 298 IRGQDNKTALYWAVEKGNATWDRDIILOCPDPTICTKOGETPLIKATKRNIEVBELLDD 357
Db 492 ARAREEQTPHIAISRLGKTEIVQLLQHMAHPDAATTNGYTPHLHISAREGQVDVASVLE 551
Qy 358 KGAKVSAVDKGDTPHLHIAIRGRSRKLAELLRL-----NPKD 394
Db 552 AGAHSATKKGFTPLHVAKYGSLDVAKLLQRRRAADSAGKNGLTPLHVAHYDNQKV 611
Qy 395 GRLL-----VRPNKAGETPNYIDC-SHOKSILTOI--FGAR-----HLSPT 432
Db 612 ALLLEKASPHATAKNGYTPHLHIAAKKQMQIASTLLNVGAETNIVTKQGVTPHLASQ 671
Qy 433 ETDGDM-----GYDLYSALADIILSEPTMQPPICVGLYAQWGSKSFLLKLEDEM-- 484
Db 672 EGHDTMVTLLDKGANIHMSTKSGLTSL-----LHLAAQ-----EDKVV 710
Qy 485 -----KTEFAGQOIEPLFQSWLIVLTLLC--GGGLL-FAFTVHPNLGIAVLSFLAL 536
Db 711 ADILTKHGADQDAHTKGLGYTPHIV-----ACHYGVNMVNFLLKQGANNAKTNGYTP 765
Qy 537 -----LYIFPIVYFGRREGESENMWAVLSTRLARHIGYLELL--LKL----- 579
Db 766 HQAAQOQHTHIINVLLQHGAKPNATTANGNTALA--IAKELGYISVVDTLKVVTEBVT 823
Qy 580 -----FVNPPPLPQGTYKALP-VRFL-P 600
Db 824 TTTITEKHLNVPEITVEVLVDSDEGDDTMTGDGGEYLRPELDELKELGDDSLPSSQFLDG 883
Qy 601 TDYRLSSVGGT-SI-----AEMIATLSDACERE 629
Db 884 MNLYRLSLEGGREDSLRSPSSDRSHYLSHASYLRDSAVMDDSVVIFSHQVSTLAKAERN 943
Qy 630 FGFLATRLFRVFKTEDTOGKKWKCTCLPSFVIFLF-----IIGCIISGITLLAIFRV 683
Db 944 SYRLS-----NGTENLDNVALLSSSPIHSGFLVIFMVDARGGAMRGCHNGLRII--I 993
Qy 684 DPKHLTVNAVLSIASVGLAFVNLNCRTWQVLDLSLLNSQRKL-----HNAASKL 734
Db 994 PPRKCTAPT-----RVTCRL-----VKRHLATWPMVEGEGLASRL 1030
Qy 735 HKLKSEG-----EMKVLKCEVELMARMAKTIDSFQNTQ-----QRLVVIIDG 776
Db 1031 IEVGPSSGAOFLGPVIVEIHFHAALRGKRELVLVRSENGSKWKEHCDYTEDSELNELLNG 1090
Qy 777 LDACEQDKVLQMLDTRVRLFSKGPFFIAIFASDPH--IIIKAINQNLNSVLNRDSNINGHDY 834
Db 1091 MD-----EVLDSPEDELEKKRICRIITRDPQYFAVVSRIKQ-----DSNLIGPE- 1134
Qy 835 MRNIVHLPVLNLSRGUSNARKFLVTSATNGDVPSCDTTGIQEDADRRVQNSLIGEMTK-- 892
Db 1135 -----GGVLSSTVVPQVQAVPEGALTKRI-----RVGLQ-----AQPMSHSELVKKI 1176
Qy 893 LGSKTALNRDITY--RRRQQRITITQMSFDLTKLAVTEDWESDISPQTMRLNIVSVT 950
Db 1177 LGNKATFSPIVLEPRRPHKFKPITWIPVPKASSDVMLNGFGGDAP-TLRL--CSIT 1232
Qy 951 GRLLRANQISFNWDRLASMINLT---EQWPYRTS-----WLILYBETEGIPQDMTLKT- 1001
Db 1233 GGTTPA-----QWEDITGITPLTFVNECVSFTTNVSARFWLI-----DCRQIQESVTFASQ 1283

Qy 1002 IYERI-----SKNIPTTKDVEPLL-----EIDGDIRNPFVFLSSRTPVLVARD 1044
Db 1284 VYREIICVPMYAKFVFAKSHDPIEARLRCFCMTDDKVDKTLQEQNF--AEVARSRD 1339
Qy 1045 VKV-----FLPCTVNLDP-----KLREIITAD-----VRAAR 1070
Db 1340 VEVLKGPPIYVDCFGNVLPLTKSGQHIFSPFAFKENRLPLFVKVDRDTQEPGCRISPFM 1399
Qy 1071 EQISIGLAYP-----ELPLHEGPPRAPSGVSYQPPSCVSTSFNGFPAGGVVSPQPHS 1123
Db 1400 EPKSTRGLVHQAICNLNITLPIY--TKESQSDQEQEEDIMTSEKN-----DETES 1448
Qy 1124 SYSGMTGQHPFPYNRGSPAPGVPVLLNSLAVDAVEKLAQIEGLDQSMLOPYCTTIKK 1183
Db 1449 TETSVL--KSH-----LVNEVPVLASPDLLSEVSEMKQDLIKM--TAILT 1489
Qy 1184 ANINGRVLAQCNIDELKEMNMNFGDWHLFRSTVLMRMAESHVVPEDPRFLSE-SSSGP 1242
Db 1490 TVUSDKA-GSIKVKELVKAABEPEG-----PREIVER-----VKEDLEKVNELRS 1337
Qy 1243 APHGEPPARASHNELPHTELSSQTPYTLNFSPEELNTLGLDEGAPRHNSLWSQSTRTP 1302
Db 1538 CTRDESSVQSSRSRSEGLVE--BEWIVSDEIEEAR-----OKAPLEITEYPCVEVRIDK 1590
Qy 1303 SLSSLSNQSSEIEISKLTQVQAEYRDAYREYIAQMSQLEGGPGSTTISGRSSPHSYTM 1362
Db 1591 EIKGRVKEKSTGLVNYLTDDLNT-CVPLPKEQLQTVODKAGKKCEALAVGRSS----- 1642
Qy 1363 QGSSSGGSIHNLQEBKGDSEB-----KPDGGRKSLMKRGDVIDYSSSGVSTN 1412
Db 1643 -----EKEGKDIPDETQSTQKHKPSLGIKKPVRRK----- 1674
Qy 1413 DASPLDPIITEDEKSDQSGSKLLPQKKSSERSLFTDLKLKSGSLRYKQLPSDEDESGT 1472
Db 1675 ----LKEQKQKEGLQASAEKAEKKGSSSESLGE-----DPGLAPEPLPTVKATS-- 1722
Qy 1473 ERSNDTPLKDDKDKAEKGVKVERVKSPE-----HSAEDIRFIKAEVYLSDALLOKDS 1527
Db 1723 -----PLIETPIGSKDKVAKLQKRVEDEQKGRSKLPFR--VKGEDVPKKTTHRPH 1774
Qy 1528 SDGVRSSB-----SSPNHSLHNEVADDSQLEKANLIELEDDSHSGKRGIPHSLSGLQD 1581
Db 1775 AASPLKSRHAPGSPSPKTERHSTLSSAKTERHPPVS-----PSSKTEKHS 1822
Qy 1582 PIARMSICEPKSPSECSLIASSPEENWPACQKAYNLNRTPTSTVTLNNSNAPANRANQ 1641
Db 1823 PV---SPSAKTERHSPASS--SKTEKSPV-----SPSTKTERHSPVSTKTER 1867
Qy 1642 NDEMEGIRETSQVILRPSSSPNPTTIQENLKS-SMTHKRSQBSYTRLSKDDPP 1694
Db 1868 H----PPVSPSGKTDKRPVSPSPGRTEKHPVSPGRTEKRLPVSPSGRTDKHQP 1917

RESULT 15

US-09-172-977-4
; Sequence 4, Application US/09172977
; Patent No. 5989863
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Guebler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Yue, Henry
; TITLE OF INVENTION: HUMAN ANKYRIN FAMILY PROTEIN
; FILE REFERENCE: PF-0615 US
; CURRENT APPLICATION NUMBER: US/09/172,977
; CURRENT FILING DATE: 1998-10-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 1839
; TYPE: PRT
; ORGANISM: Homo sapiens

Db 1774 GHTVVKVTRKIIIRYVSSSEGTKEEIMVQGMPEPNIIEGDSKVIKRVVLKSDTEQ 1833

Qy 1690 SKDPPE 1695

Db 1834 SEDNNE 1839

Search completed: February 5, 2005, 19:36:15
Job time : 70.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2005, 19:28:23 ; Search time 141 Seconds
(without alignments)
3961.558 Million cell updates/sec

Title: US-10-021-571-4
Perfect score: 8884
Sequence: 1 MSVLISQSVINYVEENIPA.....LHAAASESTGFGBERESIL 1715

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1373511 seqs, 325702437 residues

Total number of hits satisfying chosen parameters: 1373511

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8884	100.0	1715	14	US-10-021-571-4
2	8884	100.0	1715	14	US-10-117-229-2
3	8884	100.0	1715	16	US-10-473-574-26
4	8830	99.4	1771	14	US-10-117-229-4
5	8830	99.4	1771	14	US-10-117-229-11
6	8362.5	94.1	1762	14	US-10-117-229-7
7	8266	93.0	1715	14	US-10-021-571-2
8	8266	93.0	1715	14	US-10-117-229-6
9	5899.5	66.4	1184	14	US-10-117-229-3
10	5899.5	66.4	1184	14	US-10-117-229-9
11	3050.5	34.3	705	14	US-10-106-698-6378
12	2851.5	32.1	551	9	US-09-835-788A-17
13	2851.5	32.1	551	14	US-10-175-042-17

14	2638	29.7	513	14	US-10-149-819-9	Sequence 9, Appli
15	2379.5	26.8	1498	14	US-10-021-571-8	Sequence 8, Appli
16	1832.5	20.6	1398	14	US-10-021-571-6	Sequence 6, Appli
17	553	6.2	109	10	US-09-986-480-371	Sequence 371, App
18	543.5	6.1	657	15	US-10-104-047-2529	Sequence 2529, Ap
19	519	5.8	1724	9	US-09-964-899-43	Sequence 43, Appli
20	506	5.7	1330	15	US-10-108-260A-3237	Sequence 3237, Ap
21	474	5.3	1762	14	US-10-205-194-117	Sequence 117, App
22	466.5	5.3	1188	16	US-10-408-765A-1311	Sequence 1311, Ap
23	448	5.0	747	15	US-10-094-749-1924	Sequence 1924, Ap
24	444.5	5.0	919	15	US-10-108-260A-4122	Sequence 4122, Ap
25	443.5	5.0	3913	15	US-10-334-143-45	Sequence 45, Appli
26	438.5	4.9	1053	15	US-10-291-172-343	Sequence 343, App
27	438.5	4.9	1053	15	US-10-221-278-343	Sequence 343, App
28	437	4.9	740	9	US-09-835-788A-12	Sequence 12, Appli
29	437	4.9	740	14	US-10-175-042-12	Sequence 12, Appli
30	437	4.9	1569	15	US-10-275-595A-31	Sequence 31, Appli
31	409.5	4.6	1431	15	US-10-352-684A-38	Sequence 38, Appli
32	409.5	4.6	1431	15	US-10-045-400C-5	Sequence 5, Appli
33	409.5	4.6	1431	15	US-10-295-027-1235	Sequence 1235, Ap
34	409	4.6	367	15	US-10-250-613-6	Sequence 6, Appli
35	408.5	4.6	426	9	US-09-908-711-70	Sequence 70, Appli
36	395.5	4.5	1360	16	US-10-473-574-12	Sequence 12, Appli
37	390	4.4	1554	14	US-10-301-822-93	Sequence 93, Appli
38	390	4.4	1697	15	US-10-112-944-370	Sequence 370, App
39	388	4.4	622	15	US-10-112-944-345	Sequence 345, App
40	385.5	4.3	765	14	US-10-128-174-3	Sequence 3, Appli
41	385.5	4.3	765	14	US-10-128-174-34	Sequence 34, Appli
42	385.5	4.3	765	14	US-10-128-174-35	Sequence 35, Appli
43	385.5	4.3	765	14	US-10-128-174-36	Sequence 36, Appli
44	385.5	4.3	765	14	US-10-128-174-37	Sequence 37, Appli
45	385.5	4.3	765	14	US-10-128-174-38	Sequence 38, Appli

ALIGNMENTS

RESULT 1
US-10-021-571-4
; Sequence 4, Application US/10021571
; Publication No. US20030166056A1
; GENERAL INFORMATION:
; APPLICANT: CHAO, Moses V.
; APPLICANT: KONG, Haeyoung
; TITLE OF INVENTION: A TRANSMEMBRANE PROTEIN AS A DOWNSTREAM TARGET OF NEUROTROPHIN AN
; TITLE OF INVENTION: RECEPTOR TYROSINE KINASES, DNA ENCODING SAME AND MONOCLONAL ANTI
; FILE REFERENCE: CHAO11A
; CURRENT APPLICATION NUMBER: US/10/021.571
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256,909
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1715
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-021-571-4

Query Match	100.0%;	Score 8884;	DB 14;	Length 1715;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1715;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MSVLISQSVINYVEENIPALKALLEKCKVDNERNECGQTPMTAAEQGNLEIVKELIKN	60	
Db	1	MSVLISQSVINYVEENIPALKALLEKCKVDNERNECGQTPMTAAEQGNLEIVKELIKN	60	
Qy	61	GANCNLEDLONWTALISASKEGHVHVEELLKCGVNLEHRDMGWTALMWAICYKGRDVV	120	
Db	61	GANCNLEDLONWTALISASKEGHVHVEELLKCGVNLEHRDMGWTALMWAICYKGRDVV	120	
Qy	121	ELLLSHGANSVTLQYSVYPIIWAAGRGHADIYVHLLQNGAKVNCSDKYTTPLWAAAR	180	

Db 121 ELLLSHGANSVTLGQVSVVPIIWAAGRHADIVHLLQNGAKVNCSDKYGTTPLVWAAR 180
Qy 181 KGHLECVKHLHLLANGADVQEGANSMTALIVAVKGGYTQSVEILKENPNVNLTDKGNTA 240
Db 181 KGHLECVKHLHLLANGADVQEGANSMTALIVAVKGGYTQSVEILKENPNVNLTDKGNTA 240
Qy 241 LMIASKEGHEITEIVQDLDLAGTYVNIIPDRSGDTVLICAVRGHVEIVRALLQKVDIDIRG 300
Db 241 LMIASKEGHEITEIVQDLDLAGTYVNIIPDRSGDTVLICAVRGHVEIVRALLQKVDIDIRG 300
Qy 301 QDNKTALYWAVERKGNATWVDILQCNPDTEICTKGETPLIKATKMRNIEVVELLDKGA 360
Db 301 QDNKTALYWAVERKGNATWVDILQCNPDTEICTKGETPLIKATKMRNIEVVELLDKGA 360
Qy 361 KVSADVKKGDTPLHIAIRGSRKLAELLRNPKDGRLLYRPNKAGETPNIDCSHOKSL 420
Db 361 KVSADVKKGDTPLHIAIRGSRKLAELLRNPKDGRLLYRPNKAGETPNIDCSHOKSL 420
Qy 421 TQIFGARHLSPETDGMGLCYDLYSSALADILSEPTMQPPICVGLYAQWGSKSFLLKKL 480
Db 421 TQIFGARHLSPETDGMGLCYDLYSSALADILSEPTMQPPICVGLYAQWGSKSFLLKKL 480
Qy 481 EDEMKTFAQOIIEPLPQFQSWLIVFLTLCCGGIGLLFAFTVHPNLGIAVLSFLALLYIF 540
Db 481 EDEMKTFAQOIIEPLPQFQSWLIVFLTLCCGGIGLLFAFTVHPNLGIAVLSFLALLYIF 540
Qy 541 FIVYFGRREGESWNAWVLSRLARHIGYLELLKLMFVNPPPELPEOTTKALPVRFLF 600
Db 541 FIVYFGRREGESWNAWVLSRLARHIGYLELLKLMFVNPPPELPEOTTKALPVRFLF 600
Qy 601 TDYNRLLSVGGETSLEAMTATISDACERFGLATRLFRVFKTEDTQGGKKWKTCCLPS 660
Db 601 TDYNRLLSVGGETSLEAMTATISDACERFGLATRLFRVFKTEDTQGGKKWKTCCLPS 660
Qy 661 FVIFLFIICIGISGITLLAIFRVDPKHLTVNAVLISIASVVGGLAFVLCRTWMQVLDL 720
Db 661 FVIFLFIICIGISGITLLAIFRVDPKHLTVNAVLISIASVVGGLAFVLCRTWMQVLDL 720
Qy 721 NSORKELHNAASKHLKSEGFMKVLCVELMARMAKTIDSTQNTQRLVLIIDGLDAC 780
Db 721 NSORKELHNAASKHLKSEGFMKVLCVELMARMAKTIDSTQNTQRLVLIIDGLDAC 780
Qy 781 EODKVLQMLDVRVLSKGFPTAIFASDPHIIKAINQNLNSVLRDSNNGHDYMNIVH 840
Db 781 EODKVLQMLDVRVLSKGFPTAIFASDPHIIKAINQNLNSVLRDSNNGHDYMNIVH 840
Qy 841 LPVFLNSRGLSNARKFLVTSATNGDVPCSDTTGIIQEDADRRVSQNSLGEWTKLGSKTALN 900
Db 841 LPVFLNSRGLSNARKFLVTSATNGDVPCSDTTGIIQEDADRRVSQNSLGEWTKLGSKTALN 900
Qy 901 RRDYRRRQWORTITRQMSFDLTKLVTEWDFSDISPQWRRLLNIVSVTGRLLRANQIS 960
Db 901 RRDYRRRQWORTITRQMSFDLTKLVTEWDFSDISPQWRRLLNIVSVTGRLLRANQIS 960
Qy 961 FNDWRLASMINLTEQMPYRTSMILVLETEGIPDQMTLKIYERISKNIPTTKOVEP 1020
Db 961 FNDWRLASMINLTEQMPYRTSMILVLETEGIPDQMTLKIYERISKNIPTTKOVEP 1020
Qy 1021 EIDGDIRNEVEFLSSRTPVLVARDVKVFLPCTVNLDPKLEIADVRAAREQISIGLAY 1080
Db 1021 EIDGDIRNEVEFLSSRTPVLVARDVKVFLPCTVNLDPKLEIADVRAAREQISIGLAY 1080
Qy 1081 PPLPLHGGPPRAPSQVSPVCSSTSFNGPAGGVVSPQPHSSYSYSGMTGQPHPPYNRG 1140
Db 1081 PPLPLHGGPPRAPSQVSPVCSSTSFNGPAGGVVSPQPHSSYSYSGMTGQPHPPYNRG 1140
Qy 1141 SGAPGPVLLNSLVNDVACEKIKOIEGLDQSMLEPOYCTTIKKANINGVLAQCNIDELK 1200
Db 1141 SGAPGPVLLNSLVNDVACEKIKOIEGLDQSMLEPOYCTTIKKANINGVLAQCNIDELK 1200
Qy 1201 KEMNNMFGDWHLFRSTVLEWRNAESHVVPEDPRFLSESSSGPAPGEPARRASHNELPHT 1260
Db 1201 KEMNNMFGDWHLFRSTVLEWRNAESHVVPEDPRFLSESSSGPAPGEPARRASHNELPHT 1260

RESULT 2

US-10-117-229-2
; Sequence 2, Application US/10117229
; Publication No. US20030190625A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc.
; TITLE OF INVENTION: Human Kidins220Pc
; FILE REFERENCE: 9U 104 R1
; CURRENT APPLICATION NUMBER: US/10/117,229
; CURRENT FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 1715
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-117-229-2

Query Match 100.0%; Score 8884; DB 14; Length 1715;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1715; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSVLISQSVINVEBENIPALKALLEKCKVDVERNECGQTPLMIAAEOQNLEIVKELIKN 60
Db 1 MSVLISQSVINVEBENIPALKALLEKCKVDVERNECGQTPLMIAAEOQNLEIVKELIKN 60
Qy 61 GANCNLEDLNNWTALISASKEGHVHIVBELLCGVNLEHRDMGGWTALMWACVKGRTDVV 120
Db 61 GANCNLEDLNNWTALISASKEGHVHIVBELLCGVNLEHRDMGGWTALMWACVKGRTDVV 120
Qy 121 ELLLSHGANSVTLGQVSVVPIIWAAGRHADIVHLLQNGAKVNCSDKYGTTPLVWAAR 180
Db 121 ELLLSHGANSVTLGQVSVVPIIWAAGRHADIVHLLQNGAKVNCSDKYGTTPLVWAAR 180
Qy 181 KGHLECVKHLHLLANGADVQEGANSMTALIVAVKGGYTQSVEILKENPNVNLTDKGNTA 240
Db 181 KGHLECVKHLHLLANGADVQEGANSMTALIVAVKGGYTQSVEILKENPNVNLTDKGNTA 240
Qy 241 LMIASKEGHEITEIVQDLDLAGTYVNIIPDRSGDTVLICAVRGHVEIVRALLQKVDIDIRG 300

Db 241 LMIASKEGHEIIVQDLIDAGTYVNIIPDRSGDTVLIGAVRGHVEIIRALLOKYADIDIRG 300
Qy 301 QDNKTALYAWVEKGNATVRDILQCNPTETICTKDGTEPLIKATQWNIIEVVELLLDKGA 360
Db 301 QDNKTALYAWVEKGNATVRDILQCNPTETICTKDGTEPLIKATQWNIIEVVELLLDKGA 360
Qy 361 KVSADVKKGGDPLHIAIRGRSRKLAELLLRNPKGRLLYRPNKAGETPYNIDCSHQKSL 420
Db 361 KVSADVKKGGDPLHIAIRGRSRKLAELLLRNPKGRLLYRPNKAGETPYNIDCSHQKSL 420
Qy 421 TOIFGARHLSPTETDGMGLYDLYSSALADILSEPTMQPPIICVGLYAGWGSKFLLKLL 480
Db 421 TOIFGARHLSPTETDGMGLYDLYSSALADILSEPTMQPPIICVGLYAGWGSKFLLKLL 480
Qy 481 EDEMKTFAGOQIEPLFQPSWLIIVFLTLCCGLGLLPAFTVHPNLGIAVSLSFLLALYIF 540
Db 481 EDEMKTFAGOQIEPLFQPSWLIIVFLTLCCGLGLLPAFTVHPNLGIAVSLSFLLALYIF 540
Qy 541 FIVYFGRRREGESNNWAWLSTRLARHIGYLELLKLMFVNPPELPQTTKALPVRPLF 600
Db 541 FIVYFGRRREGESNNWAWLSTRLARHIGYLELLKLMFVNPPELPQTTKALPVRPLF 600
Qy 601 TDYRLSVGGETSIAEMIATLSACEREFGLATRLPRVFKTEDTQKKKWKTCCLPS 660
Db 601 TDYRLSVGGETSIAEMIATLSACEREFGLATRLPRVFKTEDTQKKKWKTCCLPS 660
Qy 661 FVIFLFIIGCIISGIIITLAIIRVDPKHLTVNAVLISIASVVGLAFVLCRTWVOVLSLL 720
Db 661 FVIFLFIIGCIISGIIITLAIIRVDPKHLTVNAVLISIASVVGLAFVLCRTWVOVLSLL 720
Qy 721 NSQRKRLHNAASKLHLKSEGFMVKLCEVELMARMAKTIDSFQNTQRLVVIIDGLDAC 780
Db 721 NSQRKRLHNAASKLHLKSEGFMVKLCEVELMARMAKTIDSFQNTQRLVVIIDGLDAC 780
Qy 781 BQDKVLOMLDTRVRLFSKGPPIALPASDPHIIKAINONLSVLIRDSNINGHDYMRNIVH 840
Db 781 BQDKVLOMLDTRVRLFSKGPPIALPASDPHIIKAINONLSVLIRDSNINGHDYMRNIVH 840
Qy 841 LPVFLNSRGLSNARKFLVTSATNGDVPCSDTGTGQEDADRVSONSGEMTKLGTKALN 900
Db 841 LPVFLNSRGLSNARKFLVTSATNGDVPCSDTGTGQEDADRVSONSGEMTKLGTKALN 900
Qy 901 RRDYTRRQMQRTITROMSFDLTLLVTEDFWFSDISPOTMERLNIIVSVTGRLLRANQIS 960
Db 901 RRDYTRRQMQRTITROMSFDLTLLVTEDFWFSDISPOTMERLNIIVSVTGRLLRANQIS 960
Qy 961 FNDRLASWINLTQWQPYRTSWLILYLEETEGIPDQWTLKTIYRISKNIPTTKDVBPPL 1020
Db 961 FNDRLASWINLTQWQPYRTSWLILYLEETEGIPDQWTLKTIYRISKNIPTTKDVBPPL 1020
Qy 1021 EIDGDIRNFVFLASRTPLVAROVKPLCTVNLDPKREIIVADVRAAREQISIGGLAY 1080
Db 1021 EIDGDIRNFVFLASRTPLVAROVKPLCTVNLDPKREIIVADVRAAREQISIGGLAY 1080
Qy 1081 PPLPLHGGPPRAGSGYSPSCVSTSGFAGGVSPQPHSSYVSGMTGPQHPFNRG 1140
Db 1081 PPLPLHGGPPRAGSGYSPSCVSTSGFAGGVSPQPHSSYVSGMTGPQHPFNRG 1140
Qy 1141 SGAPAGPVLLNSLNVDAVCEKLAQIEGLDQSLPQYCTTIKINANRVLQACNIDELK 1200
Db 1141 SGAPAGPVLLNSLNVDAVCEKLAQIEGLDQSLPQYCTTIKINANRVLQACNIDELK 1200
Qy 1201 KEMMNFGDWHLFRSTVLEMRNAESHVVPEDPRFLSESSSGPAPGHPARRASHNELPHT 1260
Db 1201 KEMMNFGDWHLFRSTVLEMRNAESHVVPEDPRFLSESSSGPAPGHPARRASHNELPHT 1260
Qy 1261 ELSQTPYTLNFSFEELNTLGLDGAHNSLWSQOTRTPPSLSLNSQSSSIEISKLT 1320
Db 1261 ELSQTPYTLNFSFEELNTLGLDGAHNSLWSQOTRTPPSLSLNSQSSSIEISKLT 1320
Qy 1321 DKVQAEYRDAYREYIAQMSQLEGGPGSTTISGRSSPHSTYTMGSSGGSIHNSLEQK 1380

Db 1321 DKVQAEYRDAYREYIAQMSQLEGGPGSTTISGRSSPHSTYTMGSSGGSIHNSLEQK 1380
Qy 1381 KDSEPKPDGGRKSFIMKRGDVIDYSSSGVSTNDASPLDPIITEEBEKSQSGSKLLPGKKS 1440
Db 1381 KDSEPKPDGGRKSFIMKRGDVIDYSSSGVSTNDASPLDPIITEEBEKSQSGSKLLPGKKS 1440
Qy 1441 SERSLSFTDLKLKXGSLRYQKLPSEDESGTEESDNTPLKDDKDKRKAEGKVERVPKSP 1500
Db 1441 SERSLSFTDLKLKXGSLRYQKLPSEDESGTEESDNTPLKDDKDKRKAEGKVERVPKSP 1500
Qy 1501 EHSAPERTFTKAEYLSLDALLDKDSSDSDSVRSSESPNHSLSHNEVADDSQLEKANLIE 1560
Db 1501 EHSAPERTFTKAEYLSLDALLDKDSSDSDSVRSSESPNHSLSHNEVADDSQLEKANLIE 1560
Qy 1561 LEDDSHGKRGIPHSLSGLQDPIIARMSICEDKKSSECSLIASSPENNPACOKAYNL 1620
Db 1561 LEDDSHGKRGIPHSLSGLQDPIIARMSICEDKKSSECSLIASSPENNPACOKAYNL 1620
Qy 1621 NRTSTVTLLNNSAPANRANQNFDEMGIRETSQVILRPSSSPNPTTIONENLKSMTHKR 1680
Db 1621 NRTSTVTLLNNSAPANRANQNFDEMGIRETSQVILRPSSSPNPTTIONENLKSMTHKR 1680
Qy 1681 SQRSSYTRLSKDPPELHAAASSESTGFGEERESIL 1715
Db 1681 SQRSSYTRLSKDPPELHAAASSESTGFGEERESIL 1715

RESULT 3

US-10-473-574-26
; Sequence 26, Application US/10473574
; Publication No. US20040116670A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; HAPALIA, April J.A.;
; APPLICANT: TANG, Y. Tom; YUE, Henry;
; APPLICANT: KHAN, Farrah A.; ISON, Craig H.;
; APPLICANT: BAUGHN, Mariah R.; WARREN, Bridget A.;
; APPLICANT: DUGGAN, Brendan M.; THANGAVELU, Kavitha;
; APPLICANT: HONCHELL, Cynthia D.; AZIMZAI, Valda;
; APPLICANT: ELLIOTT, Vicki S.; BURFORD, Neil;
; APPLICANT: DING, Li; YUE, Huibin;
; APPLICANT: BECHA, Shanya; EMERLING, Brooke M.;
; APPLICANT: RICHARDSON, Thomas W.; LEE, Soo Yeun;
; APPLICANT: BANDMAN, Olga; LAL, Preeti G.;
; APPLICANT: LEE, Sally; GIETZEN, Kimberly J.;
; APPLICANT: CHAWLA, Narinder K.; GRIFFIN, Jennifer A.;
; APPLICANT: LEE, Ernestine A.; SWARNAKAR, Anita;
; APPLICANT: RING, Huijun Z.; JONES, Karen Anne
; TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0918 USN
; CURRENT APPLICATION NUMBER: US/10/473,574
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: PCT/US02/09288
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: US 60/294,451
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: US 60/291,870
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/290,518
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US 60/288,609
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/283,769
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: US 60/281,323
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/280,508
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PERL Program
; SEQ ID NO 26
; LENGTH: 1715
; TYPE: PRT
; ORGANISM: Homo sapiens

; FEATURE: ; NAME/KEY: misc feature ; OTHER INFORMATION: Incyte ID No: 2755454CDI US-10-473--574-26									
Query Match 100.0%; Score 8884; DB 16; Length 1715; Best Local Similarity 100.0%; Pred. No. 0; Matches 1715; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MSVLSISQSVNYVEENIPAL	KALLSKCDVD	ERNCCGOTPL	MAAEOGNLEIV	KEELIKN	60		
DB	1	MSVLSISQSVNYVEENIPAL	KALLSKCDVD	ERNCCGOTPL	MAAEOGNLEIV	KEELIKN	60		
QY	61	GANCNLEDLNNW	TALISASKEGHVHIV	VELLKGVNLEHRD	MGWTALWACV	KGRTDVV	120		
DB	61	GANCNLEDLNNW	TALISASKEGHVHIV	VELLKGVNLEHRD	MGWTALWACV	KGRTDVV	120		
QY	121	ELLSSHGANS	SVTGLQVSVYPII	WAAGRHADI	VHLLLQNGAKVNC	SDKYGTTP	180		
DB	121	ELLSSHGANS	SVTGLQVSVYPII	WAAGRHADI	VHLLLQNGAKVNC	SDKYGTTP	180		
QY	181	KHLECVKHL	LANGADVQEGANS	MTALIVAVKGGYTQ	SVKEILKRNPNVNL	TDKDGNTA	240		
DB	181	KHLECVKHL	LANGADVQEGANS	MTALIVAVKGGYTQ	SVKEILKRNPNVNL	TDKDGNTA	240		
QY	241	LMTASKEG	TEIVODLLDAGTY	YNIIPDRSGDT	VLICAVRGGHV	IEIVRALLQKYAD	300		
DB	241	LMTASKEG	TEIVODLLDAGTY	YNIIPDRSGDT	VLICAVRGGHV	IEIVRALLQKYAD	300		
QY	301	QDNKTALY	WAVEKGNATW	RDILQCNPD	TEICTKDG	ETPLIKATKRNIEV	ELLDDKA	360	
DB	301	QDNKTALY	WAVEKGNATW	RDILQCNPD	TEICTKDG	ETPLIKATKRNIEV	ELLDDKA	360	
QY	361	KVSADVK	GGDTPHIAIR	GRSRKLABELL	RNPDKGRL	LYRPNKAGET	PYNI	420	
DB	361	KVSADVK	GGDTPHIAIR	GRSRKLABELL	RNPDKGRL	LYRPNKAGET	PYNI	420	
QY	421	TOIFGAR	HLSP	PTDGMGLDY	YSNADIL	SEP	TMOPPICVGLYA	QSGSGS	FLKKL
DB	421	TOIFGAR	HLSP	PTDGMGLDY	YSNADIL	SEP	TMOPPICVGLYA	QSGSGS	FLKKL
QY	481	EDEMKT	FAGQ	IEBPLFQ	SWLIVFL	TLLCGGLG	LFAFTVHP	NLGI	AVSLSL
DB	481	EDEMKT	FAGQ	IEBPLFQ	SWLIVFL	TLLCGGLG	LFAFTVHP	NLGI	AVSLSL
QY	541	FIVIFG	RRREGES	WNAW	YLSRLAR	HIGYLELL	LKLMFVNP	PELPEQ	TTKALPV
DB	541	FIVIFG	RRREGES	WNAW	YLSRLAR	HIGYLELL	LKLMFVNP	PELPEQ	TTKALPV
QY	601	TDYR	LSSV	GGETS	SLAEMI	ATLSDACER	EPFLAT	RLFRV	FKTEDTQ
DB	601	TDYR	LSSV	GGETS	SLAEMI	ATLSDACER	EPFLAT	RLFRV	FKTEDTQ
QY	661	FVIFL	FI	IGCI	ISGIT	LALFRVDP	KHLTVNA	VLI	SIASVVG
DB	661	FVIFL	FI	IGCI	ISGIT	LALFRVDP	KHLTVNA	VLI	SIASVVG
QY	721	NSQK	XLHNA	ASKL	HLKLSG	EFMKVL	KCEVEL	MARMA	KTI
DB	721	NSQK	XLHNA	ASKL	HLKLSG	EFMKVL	KCEVEL	MARMA	KTI
QY	781	EQDK	V	OML	DTVR	FLSKG	PIAF	PASD	PHI
DB	781	EQDK	V	OML	DTVR	FLSKG	PIAF	PASD	PHI
QY	841	LPV	FLNS	R	SGLS	NARK	FLV	T	ATNGD
DB	841	LPV	FLNS	R	SGLS	NARK	FLV	T	ATNGD
QY	901	RRD	Y	RRR	OM	Q	RT	T	RRR
DB	901	RRD	Y	RRR	OM	Q	RT	T	RRR

961 FNDRLASINLTEQMPYRTSWLILYLEETEGIPDMTLKTIYERISKNIPTTKOVEPLL 1020
961 FNDRLASINLTEQMPYRTSWLILYLEETEGIPDMTLKTIYERISKNIPTTKOVEPLL 1020
1021 EIDGDIRNEVFLSSRTPVLVARDVKVFLPCTVNDLPKUREIIADVRAAREQISIGLAY 1080
1021 EIDGDIRNEVFLSSRTPVLVARDVKVFLPCTVNDLPKUREIIADVRAAREQISIGLAY 1080
1081 PPLPHEGPPRAPSQSVCSSTSFNGPAGGVSPQPHSSYSGMTGPOHPYNRG 1140
1081 PPLPHEGPPRAPSQSVCSSTSFNGPAGGVSPQPHSSYSGMTGPOHPYNRG 1140
1141 SGPAPGVVLLNSLNDVAVCEKLKQIEGLDQSMLPQYCTTIKKANINGVLQAQCNIDELK 1200
1141 SGPAPGVVLLNSLNDVAVCEKLKQIEGLDQSMLPQYCTTIKKANINGVLQAQCNIDELK 1200
1201 KEMNNFQDWHLFRSTVLEMRNAESHVVPEDPRFLESSESSGPHGEPARRASHNELPHT 1260
1201 KEMNNFQDWHLFRSTVLEMRNAESHVVPEDPRFLESSESSGPHGEPARRASHNELPHT 1260
1261 ELSSQTPYTLNFSFEELNTGLDEGAPRHSNLSWOSQTRTTPSLSLNSODSSIEISKLT 1320
1261 ELSSQTPYTLNFSFEELNTGLDEGAPRHSNLSWOSQTRTTPSLSLNSODSSIEISKLT 1320
1321 DKVQAEYRDAYREYIAQMSQLEGGPGSTTISGRSSPHSTTYMGQSSSGSIHNSLQKEG 1380
1321 DKVQAEYRDAYREYIAQMSQLEGGPGSTTISGRSSPHSTTYMGQSSSGSIHNSLQKEG 1380
1381 KQSEPKDPGRKSFMLMRGDVIDYSSGSGYSTNDASPLDPIITEDEKSDQSGSKLLPGKKS 1440
1381 KQSEPKDPGRKSFMLMRGDVIDYSSGSGYSTNDASPLDPIITEDEKSDQSGSKLLPGKKS 1440
1441 SERSSLFQDMLKSGSLRYOKLPDEDESCTEESDNTPLKDDKDKAEKGVVERVPKSP 1500
1441 SERSSLFQDMLKSGSLRYOKLPDEDESCTEESDNTPLKDDKDKAEKGVVERVPKSP 1500
1501 EHSAPERTIFIKAYELSDALLDKDSDSGVRSSESSPNHSHNEVADDSQLEKANLIE 1560
1501 EHSAPERTIFIKAYELSDALLDKDSDSGVRSSESSPNHSHNEVADDSQLEKANLIE 1560
1561 LEDDSSHGKRGIPHSLSGLODPIIARMSICSEBCKKSPSECSLIASSPEENWPACQKAYNL 1620
1561 LEDDSSHGKRGIPHSLSGLODPIIARMSICSEBCKKSPSECSLIASSPEENWPACQKAYNL 1620
1621 NRTPTSVTLNNSAPANRANQNFDEMEGIRETSQVILRPSSSPNPTTIONENLKSMTWTKR 1680
1621 NRTPTSVTLNNSAPANRANQNFDEMEGIRETSQVILRPSSSPNPTTIONENLKSMTWTKR 1680
1681 SQRSSYTRLSKOPPELHAAASSESTGFGEERESIL 1715
1681 SQRSSYTRLSKOPPELHAAASSESTGFGEERESIL 1715

RESULT 4
US-10-117-229-4
; Sequence 4, Application US/10117229
; Publication No. US20030190625A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc.
; TITLE OF INVENTION: Human Kidins220Pc
; FILE REFERENCE: 9U 104 R1
; CURRENT APPLICATION NUMBER: US/10/117,229
; CURRENT FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1771
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-117-229-4

Query Match 99.4%; Score 8830; DB 14; Length 1771;
Best Local Similarity 96.7%; Pred. No. 0;

Matches 1714; Conservative 0; Mismatches 0; Indels 58; Gaps 2;			
Qy	1	MSVLISQSVINYVEENIPALKALEKCKVDNERNECCQTPLMIAAEOGNLEIVKELIKN	60
Db	1	MSVLISQSVINYVEENIPALKALEKCKVDNERNECCQTPLMIAAEOGNLEIVKELIKN	60
Qy	61	GANCNLELDNWTALISAKSGHVHIVBELLKCGVNLHRDGMGTALMWAACYKGRDGV	120
Db	61	GANCNLELDNWTALISAKSGHVHIVBELLKCGVNLHRDGMGTALMWAACYKGRDGV	120
Qy	121	ELLASHGANPSTVGIOYSVYPIIWAAGRGHADIIVHLLONGAKVNCSDKYGTTPLWAAR	180
Db	121	ELLASHGANPSTVGL-YSVYPIIWAAGRGHADIIVHLLONGAKVNCSDKYGTTPLWAAR	179
Qy	181	KHLECVKHLAMGADVDOEGANSMTALIVAKGYTQSVKEILKRNPNVNLTKDGNTA	240
Db	180	KHLECVKHLAMGADVDOEGANSMTALIVAKGYTQSVKEILKRNPNVNLTKDGNTA	239
Qy	241	LMIASKEGHTIEVDLLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQYADIDIRG	300
Db	240	LMIASKEGHTIEVDLLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQYADIDIRG	299
Qy	301	QDNKTALYAVEKGNATWVRDILQCNPDTEICTKDGTEPLIKATKMRNIEVVELLLDKGA	360
Db	300	QDNKTALYAVEKGNATWVRDILQCNPDTEICTKDGTEPLIKATKMRNIEVVELLLDKGA	359
Qy	361	KVSAVDKKGDTPLHAIIRGRSKLAELLRNPKDGRLLYRNPKAGETPNYDCSHQSKIL	420
Db	360	KVSAVDKKGDTPLHAIIRGRSKLAELLRNPKDGRLLYRNPKAGETPNYDCSHQSKIL	419
Qy	421	TOIFGARHLSPTETDGMGLDYSSALADILSEPTMQPPICVGLYAGWGSFKFLKKL	480
Db	420	TOIFGARHLSPTETDGMGLDYSSALADILSEPTMQPPICVGLYAGWGSFKFLKKL	479
Qy	481	EDEMKTFAGQOIEPLFQPSWLIIVFLTLCCGLGLLFAFTVHPNLGIAVLSFLLIYIF	540
Db	480	EDEMKTFAGQOIEPLFQPSWLIIVFLTLCCGLGLLFAFTVHPNLGIAVLSFLLIYIF	539
Qy	541	FIVYFGRRGESNNWAWLSTRLARHIGYLELLKLKMFVNPPELPTTKALPVRFPLF	600
Db	540	FIVYFGRRGESNNWAWLSTRLARHIGYLELLKLKMFVNPPELPTTKALPVRFPLF	599
Qy	601	TDYNRLSVGGTSLAEMIATLSACEREFGLATRLFRVFKETDTQKKKWKTCCLPS	660
Db	600	TDYNRLSVGGTSLAEMIATLSACEREFGLATRLFRVFKETDTQKKKWKTCCLPS	659
Qy	661	FVIFLFIIGCIISGITLLAIFRVPDKHLTVNAVLIASVVGFLAVFLNCRTWQVLDL	720
Db	660	FVIFLFIIGCIISGITLLAIFRVPDKHLTVNAVLIASVVGFLAVFLNCRTWQVLDL	719
Qy	721	NSQRKRLHNAASKLHLKSEGMVKLCEVELMARMAKTIDSPQNTQRLVVIIDGLDAC	780
Db	720	NSQRKRLHNAASKLHLKSEGMVKLCEVELMARMAKTIDSPQNTQRLVVIIDGLDAC	779
Qy	781	EQDKVLQMLDTRVRLFSKGPPIAFASDPHIIIKAINQNLNSVLDRDSNINGHDYMRNIVH	840
Db	780	EQDKVLQMLDTRVRLFSKGPPIAFASDPHIIIKAINQNLNSVLDRDSNINGHDYMRNIVH	839
Qy	841	LPVFLNSRGLSNARKFLVTSATNGDVPSCDITGQEDADRVSONSLGEMTKLASKTALN	900
Db	840	LPVFLNSRGLSNARKFLVTSATNGDVPSCDITGQEDADRVSONSLGEMTKLASKTALN	899
Qy	901	RRDYRRQMQRTITROMSPDLTKLAVTEDWFSDISPOTMERLLNIVSVTGRLLRANQIS	960
Db	900	RRDYRRQMQRTITROMSPDLTKLAVTEDWFSDISPOTMERLLNIVSVTGRLLRANQIS	959
Qy	961	FNWDLASWINLTEQWPRTSWLILYLEETEGIPDQMTLKIYERISKNIPTTKDVEPLL	1020
Db	960	FNWDLASWINLTEQWPRTSWLILYLEETEGIPDQMTLKIYERISKNIPTTKDVEPLL	1019
Qy	1021	EIDGDIRNFVFLSSRTPVLDVQKFLPCTVNLDPKRLREIIADVRAAREQISIGGLAY	1080
Db	1020	EIDGDIRNFVFLSSRTPVLDVQKFLPCTVNLDPKRLREIIADVRAAREQISIGGLAY	1079

Qy	1081	PPLPLHREPPRAPSGYQPPSVCSSTSNFGPFAGGVSPQPHSSYYSQMTGPQHPFYN--	1138
Db	1080	PPLPLHREPPRAPSGYQPPSVCSSTSNFGPFAGGVSPQPHSSYYSQMTGPQHPFYNRP	1139
Qy	1139	-----RSGSP	1143
Db	1140	FFAPLYTPRYPPGSGQHLSRPSVKTSLPRDQNNGLEVIKEDAAEGLSSPTDSRSGSP	1199
Qy	1144	APGPVLLNSLVNDVAVCEKLLQIEGLDQSMPLQYCTTIKKANINGRVLAQCINDELKEM	1203
Db	1200	APGPVLLNSLVNDVAVCEKLLQIEGLDQSMPLQYCTTIKKANINGRVLAQCINDELKEM	1259
Qy	1204	NNFCDMHLFRSTVLEMRNAESHVVPDRFLSESSSGPAPHGPPARRASHNELPHTELS	1263
Db	1260	NNFCDMHLFRSTVLEMRNAESHVVPDRFLSESSSGPAPHGPPARRASHNELPHTELS	1319
Qy	1264	SQTPYTLNFSPEELNTLGLDEGAPRHSNLSWQSOTRRTTPSLSSNSQDSSIEISKLTQKV	1323
Db	1320	SQTPYTLNFSPEELNTLGLDEGAPRHSNLSWQSOTRRTTPSLSSNSQDSSIEISKLTQKV	1379
Qy	1324	QAEYRDAYREYIAQMSOLEGGPGSTTISGRSSPHSTYYMGOSSSGSHSNLEQEGKDS	1383
Db	1380	QAEYRDAYREYIAQMSOLEGGPGSTTISGRSSPHSTYYMGOSSSGSHSNLEQEGKDS	1439
Qy	1384	EPKPDGKRSFLMKRGDVIDYSSSGVSTNDASPLDPITEEDEKSDQSGSKLLPGKKSSE	1443
Db	1440	EPKPDGKRSFLMKRGDVIDYSSSGVSTNDASPLDPITEEDEKSDQSGSKLLPGKKSSE	1499
Qy	1444	SSLFQDTLKLKGSGLRYQKLPSEDESGTESDNTPLLKDDKRAEKVERVPKSPHS	1503
Db	1500	SSLFQDTLKLKGSGLRYQKLPSEDESGTESDNTPLLKDDKRAEKVERVPKSPHS	1559
Qy	1504	AEPIRTTAKAYELSDALLDKKSDSDSVRSSESPNHSNLSHNEVADDQSLEKANLIELED	1563
Db	1560	AEPIRTTAKAYELSDALLDKKSDSDSVRSSESPNHSNLSHNEVADDQSLEKANLIELED	1619
Qy	1564	DSHSGKRGIPHSLSGLQDPIIARMSICSEDKKSPSECSLIASSPEENWPACQKAYNLNRT	1623
Db	1620	DSHSGKRGIPHSLSGLQDPIIARMSICSEDKKSPSECSLIASSPEENWPACQKAYNLNRT	1679
Qy	1624	PSTVTLNNSAPANRANQNFDEMEGIRETSQVILRPSPPNPTTIQNLKSMTHKRSQR	1683
Db	1680	PSTVTLNNSAPANRANQNFDEMEGIRETSQVILRPSPPNPTTIQNLKSMTHKRSQR	1739
Qy	1684	SSYTRLKDPPELHAAASSESTGFEERESIL	1715
Db	1740	SSYTRLKDPPELHAAASSESTGFEERESIL	1771

RESULT 5
US-10-117-229-11
; Sequence 11, Application US/10117229
; Publication No. US20030190625A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc.
; TITLE OF INVENTION: Human Kidine220pc
; FILE REFERENCE: 9U 104 RI
; CURRENT APPLICATION NUMBER: US/10/117,229
; CURRENT FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1771
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-117-229-11

Query Match 99.4%; Score 8830; DB 14; Length 1771;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 1714; Conservative 0; Mismatches 0; Indels 58; Gaps 2;
Qy 1 MSVLISQSVINYVEENIPALKALEKCKVDNERNECCQTPLMIAAEOGNLEIVKELIKN 60

Db 1 MSVLISQSVINYVEENIPALKALLEKCKVDVERNECGQTPMLAAEQGNLEIVKELIKN 60
QY 61 GANCNLEDLNDWTALISASKEGHVHIVEELLKCGVNLHRDMGWTALWACYKGRDVV 120
Db 61 GANCNLEDLNDWTALISASKEGHVHIVEELLKCGVNLHRDMGWTALWACYKGRDVV 120
QY 121 ELLLSHGANSVTGLQSVYVPIIWAAGRGHADIVHLLQNGAKVNSDKYGTTPVWAAR 180
Db 121 ELLLSHGANSVTGLQSVYVPIIWAAGRGHADIVHLLQNGAKVNSDKYGTTPVWAAR 179
QY 181 KGHLECVKHLAMGADVDOGANSMTALIVAVKGGYTQSVKEILKKNPNVNLTKDGN 240
Db 180 KGHLECVKHLAMGADVDOGANSMTALIVAVKGGYTQSVKEILKKNPNVNLTKDGN 239
QY 241 LMIASKEGHTIEIVQDLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRG 300
Db 240 LMIASKEGHTIEIVQDLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRG 299
QY 301 QDNKTALYWAVERKGNATWVRDILQCNPDTEICTKGETPLIKATKWRNIEVVELLDKGA 360
Db 300 QDNKTALYWAVERKGNATWVRDILQCNPDTEICTKGETPLIKATKWRNIEVVELLDKGA 359
QY 361 KVSADVKKGTPLHIAIRGSRKLAELLNPNKDGRLLYRPNKAGETPPYINIDCSHOKSIL 420
Db 360 KVSADVKKGTPLHIAIRGSRKLAELLNPNKDGRLLYRPNKAGETPPYINIDCSHOKSIL 419
QY 421 TQIFGARHLSPTETDGMGLYDLYSSALADILSEPTMQPPICVGLYAQMGSGKSFLLKKL 480
Db 420 TQIFGARHLSPTETDGMGLYDLYSSALADILSEPTMQPPICVGLYAQMGSGKSFLLKKL 479
QY 481 EDEMKTIFAGQIIEPLQFQSWLIVFLTLICGGLGLLPAFTVHNLGIAVSLSPALLYIF 540
Db 480 EDEMKTIFAGQIIEPLQFQSWLIVFLTLICGGLGLLPAFTVHNLGIAVSLSPALLYIF 539
QY 541 FIVYFGRREGESWNAWVLSRLARHIGYLELLKLMFVNPELPEQTTKALPVRELF 600
Db 540 FIVYFGRREGESWNAWVLSRLARHIGYLELLKLMFVNPELPEQTTKALPVRELF 599
QY 601 TDYNRLSSVGGETSLAEMIATLSDACEREFGLATRLFRVFKTEDTQKKKWKTKCLPS 660
Db 600 TDYNRLSSVGGETSLAEMIATLSDACEREFGLATRLFRVFKTEDTQKKKWKTKCLPS 659
QY 661 FVIFLFIIGCIISGITLLAI PRVDPKHLTVNAVLISIASVVGGLAFVNLCKRTWQVLDL 720
Db 660 FVIFLFIIGCIISGITLLAI PRVDPKHLTVNAVLISIASVVGGLAFVNLCKRTWQVLDL 719
QY 721 NSQKRLHNAASKLHLKSGFGMKVLKCEVELMARWAKTIDSFQNTQRLVLIIDGLDAC 780
Db 720 NSQKRLHNAASKLHLKSGFGMKVLKCEVELMARWAKTIDSFQNTQRLVLIIDGLDAC 779
QY 781 EQDKVLQMLDTRVFLSKGPFIAIFASDPHIIKAINQNLSVLRDSNNGHDMYRNIVH 840
Db 780 EQDKVLQMLDTRVFLSKGPFIAIFASDPHIIKAINQNLSVLRDSNNGHDMYRNIVH 839
QY 841 LPVFLNSRGLSNARKFLVTSATNGDVPSCDTTIGQEDADRVSQNSIGEMTKLGSKTALN 900
Db 840 LPVFLNSRGLSNARKFLVTSATNGDVPSCDTTIGQEDADRVSQNSIGEMTKLGSKTALN 899
QY 901 RRDYRRORQRTITRQMSFDLTKLVTEWDFSDISQPMRRLLNTVSVTGRLLRANQIS 960
Db 900 RRDYRRORQRTITRQMSFDLTKLVTEWDFSDISQPMRRLLNTVSVTGRLLRANQIS 959
QY 961 FNNWDLASWINLQEQPYRTSWLILYLEETEGIPDQMTLKIYERISKNIPTTKDVEPLL 1020
Db 960 FNNWDLASWINLQEQPYRTSWLILYLEETEGIPDQMTLKIYERISKNIPTTKDVEPLL 1019
QY 1021 EIDGDIRNFVFLSSRTPVLVARDVKVFLPCTVNLDPKLEIITADVRAAREQISIGLAY 1080
Db 1020 EIDGDIRNFVFLSSRTPVLVARDVKVFLPCTVNLDPKLEIITADVRAAREQISIGLAY 1079
QY 1081 PPLPLHEGPPRAPSGYSQPSVCSSTSFNGPAGGVVSPQPHSSYYSGMTGPOHPFN 1138

Db 1080 PPLPLHEGPPRAPSGYSQPSVCSSTSFNGPAGGVVSPQPHSSYYSGMTGPOHPFN 1139
QY 1139 -----RGSGP 1143
Db 1140 FFAPLYTPRYYPGSGHLSRPSVKTSPLRQDNNGLEVIKEDAAEGLSPSTDSSRGSGP 1199
QY 1144 AGPVLVLSLNDVACEKLOIEGLDOSMLPOYCTTIKKANINGVLAQCNDIELKKEM 1203
Db 1200 AGPVLVLSLNDVACEKLOIEGLDOSMLPOYCTTIKKANINGVLAQCNDIELKKEM 1259
QY 1204 NMNFGDWHLFRSTVLEMRNAESHVPEDPFISESSSGPAPGEPARRASHNELPHTELS 1263
Db 1260 NMNFGDWHLFRSTVLEMRNAESHVPEDPFISESSSGPAPGEPARRASHNELPHTELS 1319
QY 1264 SOTPYTLNFPSELNLTGLDEGAPRHSNLSWQSOTRTPSLSLNSQDSSEISKLTDKV 1323
Db 1320 SOTPYTLNFPSELNLTGLDEGAPRHSNLSWQSOTRTPSLSLNSQDSSEISKLTDKV 1379
QY 1324 QAEYRDAYREYTAQNSOLEGGPGSTTISGRSPHSTVYMGSSSGSIHNSLQOEKGD 1383
Db 1380 QAEYRDAYREYTAQNSOLEGGPGSTTISGRSPHSTVYMGSSSGSIHNSLQOEKGD 1439
QY 1384 EPKPDGRKSFMLKRGDVIDYSSGVSTNDASPLDPIITEDEKSDQSGKLLPGKKSSER 1443
Db 1440 EPKPDGRKSFMLKRGDVIDYSSGVSTNDASPLDPIITEDEKSDQSGKLLPGKKSSER 1499
QY 1444 SSLFQTDLKLKSGGLRYQKLPDEDESGTEESDNTPLKDDKDKAEGKVERVPKSPHS 1503
Db 1500 SSLFQTDLKLKSGGLRYQKLPDEDESGTEESDNTPLKDDKDKAEGKVERVPKSPHS 1559
QY 1504 APPIRTFIKAYEVLSDALLDKKSDSGVRSSESSNHSILHNEVADDSOLEKANLIELED 1563
Db 1560 APPIRTFIKAYEVLSDALLDKKSDSGVRSSESSNHSILHNEVADDSOLEKANLIELED 1619
QY 1564 DSHSGRGIPIHSLSGLODPIIARMSICSEDKSPSCSLIASSPENWPAKOKAYNLNRT 1623
Db 1620 DSHSGRGIPIHSLSGLODPIIARMSICSEDKSPSCSLIASSPENWPAKOKAYNLNRT 1679
QY 1624 PSTVTLNNSAPANRANQNFDEMEGIRETSQVILRPSSESPNPTTIQENLKSMTKRQR 1683
Db 1680 PSTVTLNNSAPANRANQNFDEMEGIRETSQVILRPSSESPNPTTIQENLKSMTKRQR 1739
QY 1684 SSTRLSKDPPELHAAASSESTGFGBERESIL 1715
Db 1740 SSTRLSKDPPELHAAASSESTGFGBERESIL 1771

RESULT 6

US-10-117-229-7
; Sequence 7, Application US/10117229
; Publication No. US20030190625A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies, Inc.
; TITLE OF INVENTION: Human Kidins220Pc
; FILE REFERENCE: 9U 104 R1
; CURRENT APPLICATION NUMBER: US/10/117,229
; CURRENT FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1762
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-117-229-7

Query Match 94.1%; Score 8362.5; DB 14; Length 1762;
Best Local Similarity 91.6%; Pred. No. 0;
Matches 1616; Conservative 47; Mismatches 50; Indels 51; Gaps 4;
QY 1 MSVLISQSVINYVEENIPALKALLEKCKVDVERNECGQTPMLAAEQGNLEIVKELIKN 60
Db 1 MSVLISQSVINYVEENIPALKALLEKCKVDVERNECGQTPMLAAEQGNLEIVKELIKN 60

QY 61 GANCNLELDLNNWTALISAKSGHVHIVBELLKCGVNLHRDMGGWTALMWACYKGRTDV 120
DB 61 GANCNLELDLNNWTALISAKSGHIIHVBELLKSGASLHRDMGGWTALMWACYKGRTDV 120
QY 121 ELLLSHGANSVTCLOYSVYPIIWAAGRGHADIHLLLLONGAKVNCSDKYGTTPVWAAR 180
DB 121 ELLLSHGANSVTCGL-YSVYPIIWAAGRGHADIHLLLLONGAKVNCSDKYGTTPVWAAR 179
QY 181 KGHLECVKHLAMGADVDEGANSMTALIVAKGGYTQSVKBIKRNPNVNLTKDGN 240
DB 180 KGHLECVKHLAMGADVDEGANSMTALIVAKGGYTQSVKBIKRNPNVNLTKDGN 239
QY 241 LMIAKSGHTEIVQDLLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQYADIDIRG 300
DB 240 LMIAKSGHTEIVQDLLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQYADIDIRG 299
QY 301 QDNKTALYWAVEKGNATVMDILOCNPDTEICTKDGTEPLIKATMNRNIEVVELLDKGA 360
DB 300 QDNKTALYWAVEKGNATVMDILOCNPDTEICTKDGTEPLIKATMNRNIEVVELLDKGA 359
QY 361 KVSADVKKGDTPLHIAIRGRSKLAELLLRNPDKGRLLYRNKAGETPNYIDCSHOKSIL 420
DB 360 KVSADVKKGDTPLHVAIRGRSRLAELLLRNPDKGRLLYRNKAGETPNYIDCSHOKSIL 419
QY 421 TQIFGARHLSPTETDGMGLYDLYSSALADILSBPTMOPPICVGLYAOQWGSFKLLK 480
DB 420 TQIFGARHLSPTETDGMGLYDLYSSALADILSBPTMOPPICVGLYAOQWGSFKLLK 479
QY 481 EDEMKTFAGQIEPLFQPSWLIIVFLTLCCGLGLLFAFTVHPLNGIAVLSPLALLYIP 540
DB 480 EDEMKTFAGQTEPLFQPSWLIIVFLTLCCGLGLLFAFTVDTNLAIASLSPLALIYIF 539
QY 541 FIVYFGRGREGESNNWAWLSTLARHIGYLELLKLMLFVNPPELPOTTKALPVRF 600
DB 540 FIVYFGRGREGESNNWAWLSTLARHIGYLELLKLMLFVNPPELPOTTKALPVRF 599
QY 601 TDYNRLSVGETSLAEMIATLSACEREFGLATRLFRVFKTEDTQCKKKWKTCC 660
DB 600 TDYNRLSVGETSLAEMIATLSACEREFGLATRLFRVFKTEESQCKKKWKTCC 659
QY 661 FVIFLFIIGCIISGITLLAIIRVPDKHLTVNAVLIASVVGFLAVLNCRTWQVLSLL 720
DB 660 FVIFLFIIGCIISGITLLAIIRVPDKHLTVNAVLIASVVGFLAVLNCRTWQVLSLL 719
QY 721 NSQKRLHNAASKLHLKSEGFVKLVCEVELMARMKTIIDFTONOTRLVVIIDGLDAC 780
DB 720 NSQKRLHSAASKLHLKSEGFVKLVCEVELMARMKTIIDFTONOTRLVVIIDGLDAC 779
QY 781 EQDKVLQMLDTRVRLFSKGPPIAFASDPHIIKAINQNLNSVLRLDSNININGHDYMRNIVH 840
DB 780 EQDKVLQMLDTRVRLFSKGPPIAFASDPHIIKAINQNLNSVLRLDSNININGHDYMRNIVH 839
QY 841 LPVFLNSRGLSNARKFLVTSATNGDVPDSTDTGQEDADRRVSONSGEMTKLGSKTALN 900
DB 840 LPVFLNSRGLSNARKFLVTSATNGDITCSDTGTGQEDTDRVSONSGEMTKLGSKTALN 899
QY 901 RRDYVRQORTITROMSFDLTKLAVTEWFSDISPOTMERLANIVSVTCRLRANOIS 960
DB 900 RRDYVRQORTITROMSFDLTKLAVTEWFSDISPOTMRRLNIVSVTCRLRANOIT 959
QY 961 FNWDLASWINLTQWPYRTSWLLIYLEETEGIPDQMTLKIYERISKNIPTTKDVEP 1020
DB 960 FNWDLASWINLTQWPYRTSWLLIYLEETEGIPDQMTLKIYERISKNIPTTKDVEP 1019
QY 1021 EIDGDIRNFVFLSSRTPVLVARDVKVFLPCTVNLDPKLRHIIADVRAAREIQSIGGLAY 1080
DB 1020 EIDGDIRNFVFLSSRTPVLVARDVKVFLPCTVNLDPKLRHIIADVRAAREIQINIGGLAY 1079
QY 1081 PPLPLHEGPPRPPSGYSPQPSVCSSTFNGPFAGVUSPOPHSSVYSGMTGPQHPFYNR- 1139
DB 1080 PPLPLHEGPPRPPSGYSPQPSVCSSTFNGPFAGVUSPOPHSSVYSGMTGPQHPFYNR 1139
QY 1140 -----GSG-----PAGPVL 1151

DB 1140 FFAPVLYTPRYPGSQHLISRSSVKTSLSLPRDQNNGLPCDSGFKNQQAAPATGSSLL 1199
QY 1152 NSLVNDVACEKLOIEGLDQSMLOPYCTTIKANKINGVLAOCNIDELKEMAMNFGDWH 1211
DB 1200 SSMTVDVCEKLOIEGLDQSMMPQYCTTIKANKINGVLSQCNIDELKEMAMNFGDWH 1259
QY 1212 LFRSTVLEMRNAESHVVPEDPRFLSESSSGPAPGEPARRASHNELPHELTELSSQTPYTLN 1271
DB 1260 LFRSVLEMRSVESQVVPEDPRFLNENSSAPVPHGESARRSHTELPTELSSQTPYTLN 1319
QY 1272 PSFELMTGLDEGAPRHSNLSWQSOTRRTPLSLNSQDSSIEISKLTDKVQAEYRDAY 1331
DB 1320 PSFELMTGLDEGAPRHSNLSWQSOTRRTPLSLNSQDSSIEISKLTDKVQAEYRDAY 1379
QY 1332 REYTAQMSOLEGGPGSTTISGRSSPHSTYYMGQSSSGSIHSNLEQEKGDSEPKPDGR 1391
DB 1380 REYTAQMSOLEGGTGSSTISGRSSPHSTYYIGOSSSGSIHSTLEQERKGEGLKQEDGR 1439
QY 1392 KSFLMKRGDVIDYSSGVSTNDASPLDPIITEDEKSDQSGSKLLPGKSSRSRSLFQTDL 1451
DB 1440 KSFLMKRGDVIDYSSGVSTNEASPLDPIITEDEKSDQSGSKLLPGKSSRSRSLFQTDL 1499
QY 1452 KLKSGGLRYQKLPSEDESGTEESDNTPLKDKDKRKAEGKVERVPKSPHSABPIRTFI 1511
DB 1500 KLKSGGLRYQKLPSEDESGTEESDNTPLKDKDKRKAEGKVERVPKSPHSABPIRTFI 1559
QY 1512 KAKELYSDALLDKKSDSDSGVRSSSESPHSNLSHNEVADDSOLEKANLIELEDDSHSGKRG 1571
DB 1560 KAKELYSDALLDKKSDSDSGVRSSSESPHSNLSHNEVADDSOLEKANLIELEDDSHSGKRG 1619
QY 1572 IPHLSGLQDPIIARMSICSEDKKSPSECSLIASSPEENWPACOKAYNLNRTPTVTLLN 1631
DB 1620 MPHLSGLQDPIIARMSICSEDKKSPSECSLIASSPEESWPACOKAYNLNRTPTVTLLN 1679
QY 1632 NSAPANRANQNFDMEGIRETSQVILRPSSSPNPTTIQENLKMTHKRSQSSYTRLSK 1691
DB 1680 NTAPTNANQNFDEIGIRETSQVILRPSPNPTAQVNENLKMMAHKRSQSSYTRLSK 1739
QY 1692 DPPELHAAASESTGFGEERESIL 1715
DB 1740 DASELH-AASESTGFGEERESIL 1762

RESULT 7

US-10-021-571-2
; Sequence 2, Application US/10021571
; Publication No. US20030166056A1
; GENERAL INFORMATION:
; APPLICANT: KONG, Haeyoung
; TITLE OF INVENTION: A TRANSMEMBRANE PROTEIN AS A DOWNSTREAM TARGET OF NEUROTROPHIN AN
; TITLE OF INVENTION: RECEPTOR TYROSINE KINASES, DNA ENCODING SAME AND MONOCLONAL ANTI
; FILE REFERENCE: CHAO11A
; CURRENT APPLICATION NUMBER: US/10/021,571
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256,909
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1715
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-021-571-2

Query Match 93.0%; Score 8266; DB 14; Length 1715;
Best Local Similarity 92.7%; Pred. No. 0;
Matches 1591; Conservative 53; Mismatches 70; Indels 2; Gaps 2;
QY 1 MSLVISQSVINYVEENIPALKLEKCKVDNERNECGQTPMLTAASQGNLEIVKELIKN 60
DB 1 MSLVISQSVINYVEENIPALKLEKCKVDNERNECGQTPMLTAASQGNVEIVKELIKN 60

```
QY 61 GANCNLEDDNWTALISASKEGHVHIVEELLKCGVNLHHRDMGWTALMWACYKGRDVV 120
Db 61 GANCNLEDDNWTALISASKEGHVHIVEELLKCGVNLHHRDMGWTALMWACYKGRDVV 120
QY 121 ELLSHGANPSVTGLQVSVYPIIWAAGRHADIVHLLLLQNGAKVNSDKYGTTPLVAAAR 180
Db 121 ELLSHGANPSVTGLQVSVYPIIWAAGRHADIVHLLLLQNGAKVNSDKYGTTPLVAAAR 180
QY 181 KGHLECVKHLAMGADVQEGANSMTALIIVAVRGYVTSQVKEILKRNPNVNLTKDGNNTA 240
Db 181 KGHLECVKHLAMGADVQEGANSMTALIIVAVRGYVTSQVKEILKRNPNVNLTKDGNNTA 240
QY 241 LMTASKEGHEIIVQDLDAGTYNIPDRSGDTVLIGAVRGHVEIVRALLQKADIDIRG 300
Db 241 LMTASKEGHEIIVQDLDAGTYNIPDRSGDTVLIGAVRGHVEIVRALLQKADIDIRG 300
QY 301 QDNKTALYAVEKGNATWVRDILQCNPDTEICTKDGETPLIKATKRNIEVVELLDKGA 360
Db 301 QDNKTALYAVEKGNATWVRDILQCNPDTEICTKDGETPLIKATKRNIEVVELLDKGA 360
QY 361 KVSADVKKGDTPLHVAIRGRSRRLAELLRNPKDGRLLYRPNKAGETPPYINDCSHQKSL 420
Db 361 KVSADVKKGDTPLHVAIRGRSRRLAELLRNPKDGRLLYRPNKAGETPPYINDCSHQKSL 420
QY 421 TQIFGARHLSPTEITDGMGLDYDLYSSALADILSEPTMOPPICVGLYAQWGSFKLLKKL 480
Db 421 TQIFGARHLSPTEITDGMGLDYDLYSSALADILSEPTMOPPICVGLYAQWGSFKLLKKL 480
QY 481 EDEWKTFAQOIQBPLOFQSWLIIVFLTLCCGGLGLFAFVPHNLGTAVALSLFALLIYF 540
Db 481 EDEWKTFAQOIQBPLOFQSWLIIVFLTLCCGGLGLFAFVPHNLGTAVALSLFALLIYF 540
QY 541 FIVIFGGRREGESWNMAWVLSRLARHIGYLELLKLMFVNPPPELPEQTTKALPVRFLF 600
Db 541 FIVIFGGRREGESWNMAWVLSRLARHIGYLELLKLMFVNPPPELPEQTTKALPVRFLF 600
QY 601 TDYNRSLSSVGGETSIAEMIATLSDACERERGFATRLFRVFKTEDTQKKKKWKTCLPS 660
Db 601 TDYNRSLSSVGGETSIAEMIATLSDACERERGFATRLFRVFKTEDTQKKKKWKTCLPS 660
QY 661 FVIFLFIIGCIISGITLTAIFRVDPKHLTVNAVLIISIASVVGFLAVLNCRTHWQVLDL 720
Db 661 FVIFLFIIGCIISGITLTAIFRVDPKHLTVNAVLIISIASVVGFLAVLNCRTHWQVLDL 720
QY 721 NSQKRLHNAASKLHLKLSGFMKVLKCEVELMARMAKTIDSFTQNTQLRWIIDGLDAC 780
Db 721 NSQKRLHNAASKLHLKLSGFMKVLKCEVELMARMAKTIDSFTQNTQLRWIIDGLDAC 780
QY 781 EQDKVLQMLDTRVLFSGKGFPIAFASDPHIIKAINQNLNVLNDSNINCHDYMNRIVH 840
Db 781 EQDKVLQMLDTRVLFSGKGFPIAFASDPHIIKAINQNLNVLNDSNINCHDYMNRIVH 840
QY 841 LPVFLNSRGLSNARKFLVTSATNGDVPCTDTGQEDADRRVQNSIGEMTKLGSKTALN 900
Db 841 LPVFLNSRGLSNARKFLVTSATNGDVPCTDTGQEDADRRVQNSIGEMTKLGSKTALN 900
QY 901 RRDYTRRQMQRTITRQMSFDLTKLVTEDFSDISPTQMRRLNIVSVTGRLLRANQIS 960
Db 901 RRDYTRRQMQRTITRQMSFDLTKLVTEDFSDISPTQMRRLNIVSVTGRLLRANQIT 960
QY 961 FNMDRLASWINLTQWQPYRTSWLLIYLBETEGIPDQMTLKIYERISKNIPPTKDVPEPL 1020
Db 961 FNMDRLASWINLTQWQPYRTSWLLIYLBETEGIPDQMTLKIYERISKNIPPTKDVPEPL 1020
QY 1021 EIDGDIENFEVLSRTPVLVARDVKVFLPCTVNLDPKLEIIVADVRAAREQISIGLAY 1080
Db 1021 EIDGDIENFEVLSRTPVLVARDVKVFLPCTVNLDPKLEIIVADVRAAREQINIGLAY 1080
QY 1081 PPLPLHEGPPRAPSGYQPPSVCSSTSFNGFPAGVVSVPQPHSGYSYSGMTGPQHPFYNRG 1140
Db 1081 PPLPLHEGPPRAPSGYQPPSVCSSTSFNGFPAGVVSVPQPHSGYSYSGMTGPQHPFYNRG 1140
```

```
QY 1141 SGPAPGPVVLLNSLNDVAVCEKQLQIEGLDQSMLOYCTTIKKANINGRVLACQNDIELK 1200
Db 1141 AVPATGSSLLLSGSMVTVDVCEKRLQIEGLDQSMLOYCTTIKKANINGRVLACQNDIELK 1200
QY 1201 KEMNNMFGDWHILFRSTVLEMRNAESHVVPEDPRFLSESSSGPAPGEPARRASHNELPHT 1260
Db 1201 KEMANNMFGDWHILFRSMVLEMRSVESQVVPEDPRFLNENSSAPVPHGESARRSHTPLT 1260
QY 1261 ELSQTPYTLNFSPELNTLGLDEGAPRHSNLSWQSTRTTSLSLNSQDSSIEISKL 1320
Db 1261 ELSQTPYTLNFSPELNTLGLDEGAPRHSNLSWQSTRTTSLSLNSQDSSIEISKL 1320
QY 1321 DKVQAYRDAYREYIAQMSOLEGPGSTTISGRSSPHSTVYMGQSSGSGSIHNLQEK 1380
Db 1321 DKVQAYRDAYREYIAQMSOLEGPGSTTISGRSSPHSTVYMGQSSGSGSIHNLQEK 1380
QY 1381 KQSEPKPDGRKSFMLKRGDVIDYSSGVSTNDASPLDPIITEDEKSDQSGSKLLPKKS 1440
Db 1381 KEGELKQEDGRKSFMLKRGDVIDYSSGVSTNEASPLDPIITEDEKSDQSGSKLLPKKS 1440
QY 1441 SERSSLFQTDLKLKSGGLRYQKLPSEDESGTSEESNTPLKDDKDKRKAEGKVERVPKS 1499
Db 1441 SERPSLFQTDLKLKGGGLRYQKLPSEDESGTGERVQITPHCSKMITKRLKAKQRCASP 1500
QY 1500 PHSASPIRTFIKAKYLSDDLKDDSDSGVRSSESPNHSNHNVEADDSOLEKANLI 1559
Db 1501 QHSASPIRTFIKAKYLSDDLKDDSDSGVRSSESPNHSNHNVEADDSOLEKANLI 1560
QY 1560 ELEDDSHSGKRGIPHSLGLQDPIIARMSICSEDKSPSECSLIASSPEENWPACQKAYN 1619
Db 1561 ELEDEGHSGKRGMPHSLGLQDPIIARMSICSEDKSPSECSLIASSPEESWPACQKAYN 1620
QY 1620 LNRTPTVTILNNSAPANRANQNFDEMEGIRETSQVILRPPSSPNPTTQNEHLKSMTHK 1679
Db 1621 LNRTPTVTILNNTAPTNRANQNFDEMEGIRETSQVILRPPSPNPTAVQNLNLSMAHK 1680
QY 1680 RSQRSSYTRLSKDPPELHAAASESTGFGEERESIL 1715
Db 1681 RSQRSSYTRLSKDSSELH-AAASESTGFGEERESIL 1715
```

RESULT 8

```
US-10-117-229-6
; Sequence 6, Application US/10117229
; Publication No. US20030190625A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc.
; TITLE OF INVENTION: Human Kidins220Pc
; FILE REFERENCE: 9U 104 R1
; CURRENT APPLICATION NUMBER: US/10/117,229
; CURRENT FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 6
; LENGTH: 1715
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; US-10-117-229-6
```

```
Query Match 93.0%; Score 8266; DB 14; Length 1715;
Best Local Similarity 92.7%; Pred. No. 0;
Matches 1591; Conservative 53; Mismatches 70; Indels 2; Gaps 2;
```

```
QY 1 MSVLISQSVINYVEENIPALKALLEKCKDVDERNECGQTPLMIAAQGNLEIVKELIKN 60
Db 1 MSVLISQSVINYVEENIPALKALLEKCKDVDERNECGQTPLMIAAQGNLEIVKELIKN 60
QY 61 GANCNLEDDNWTALISASKEGHVHIVEELLKCGVNLHHRDMGWTALMWACYKGRDVV 120
Db 61 GANCNLEDDNWTALISASKEGHVHIVEELLKCGVNLHHRDMGWTALMWACYKGRDVV 120
QY 121 ELLSHGANPSVTGLQVSVYPIIWAAGRHADIVHLLLLQNGAKVNSDKYGTTPLVAAAR 180
```

Db 121 ELLLSHGANSVTVGLQYVYPIIWAAGRGHADIHLLLLONGAKVNCSDKYGTTPVJVAAR 180
Qy 181 KGHLECVKHLAMGADVQEGANSMTALIIVAKGGYTQSVKEILKRNPNVNLTKDGNNTA 240
Db 181 KGHLECVKHLAMGADVQEGANSMTALIIVAKGGYTQSVKEILKRNPNVNLTKDGNNTA 240
Qy 241 LMIAKSGHTEIVQDLIDAGTVYNIIPDRSGDTVLIIVAGRGHVEIVRALLQYADIDIRG 300
Db 241 LMIAKSGHTEIVQDLIDAGTVYNIIPDRSGDTVLIIVAGRGHVEIVRALLQYADIDIRG 300
Qy 301 QDNKTALYAVEKGNATVVDILQCNPDTEICTKDGTEPLIKATKORNIIEVEILLDKGA 360
Db 301 QDNKTALYAVEKGNATVVDILQCNPDTEICTKDGTEPLIKATKORNIIEVEILLDKGA 360
Qy 361 KVSADVKKGDTPLHAIIRGRSKLAELLRLNPKGRILLYRNPKAGETPNYIDCSHOKSIL 420
Db 361 KVSADVKKGDTPLHAIIRGRSKLAELLRLNPKGRILLYRNPKAGETPNYIDCSHOKSIL 420
Qy 421 TOIFGARHLSPTETDGMGLYDLYSSALADILSEPTMOPPICVGLYAOWGSGKFLKKL 480
Db 421 TOIFGARHLSPTETDGMGLYDLYSSALADILSEPTMOPPICVGLYAOWGSGKFLKKL 480
Qy 481 EDEMKTPAQOIEPLFQFQSWLIVFLTLCCGLGLLPAFTVHPNIGIAVLSFLALYIF 540
Db 481 EDEMKTPAQOIEPLFQFQSWLIVFLTLCCGLGLLPAFTVHPNIGIAVLSFLALYIF 540
Qy 541 FIVYFGRRGESWNWAWLSTRLARHIGYLELLKLMFVNPPELPTQTKALPVRF 600
Db 541 FIVYFGRRGESWNWAWLSTRLARHIGYLELLKLMFVNPPELPTQTKALPVRF 600
Qy 601 TDYNRLSVGGTSLAEMIATLSACEREFGLATRLFRVPKTEDTOGKKWKTCCLPS 660
Db 601 TDYNRLSVGGTSLAEMIATLSACEREFGLATRLFRVPKTEESQKKWKTCCLPS 660
Qy 720 FVIFLFIIGCIIIGITLLAIIRVDPKHLTVNAVLISIASVVGGLAFVNLNCRVWQVLSLL 720
Db 720 FVIFLFIIGCIIIGITLLAIIRVDPKHLTVNAVLISIASVVGGLAFVNLNCRVWQVLSLL 720
Qy 721 NSQRKRLHNAASKLHLKSEGFVKLCEVELMARMATIDSFTQNTRLVVIIDGLDAC 780
Db 721 NSQRKRLHNAASKLHLKSEGFVKLCEVELMARMATIDSFTQNTRLVVIIDGLDAC 780
Qy 781 EODKVLQMLDTRVRLFSKGPPIAFASDPHIIKAINQNLNSVLDRSNGHDMYRNVH 840
Db 781 EODKVLQMLDTRVRLFSKGPPIAFASDPHIIKAINQNLNSVLDRSNGHDMYRNVH 840
Qy 841 LPVFLNSRGLSNARKFLVTSATNGDVCSTTQEDADRRVSONSLGEMTKLGSKTALN 900
Db 841 LPVFLNSRGLSNARKFLVTSATNGDVCSTTQEDADRRVSONSLGEMTKLGSKTALN 900
Qy 901 RRDYRRROMORTITROMSPDLTKLLVTEWFDSDI SPQWRLNINYSVTGRLRANQIS 960
Db 901 RRDYRRROMORTITROMSPDLTKLLVTEWFDSDI SPQWRLNINYSVTGRLRANQIT 960
Qy 961 FNWDLASWINLTQWVPRTSWLILYEETEGIPDQMTLKIYERISKNIPTTKDVEPLL 1020
Db 961 FNWDLASWINLTQWVPRTSWLILYEETEGIPDQMTLKIYERISKNIPTTKDVEPLL 1020
Qy 1021 EIDGDIRNFVFLSSRTPVLVARDVKVFLPCTVNLDPKRLREIIADVRAAREQINIGL 1080
Db 1021 EIDGDIRNFVFLSSRTPVLVARDVKVFLPCTVNLDPKRLREIIADVRAAREQINIGL 1080
Qy 1081 PPLPHGPPRPPAGSYQPPSVCSSTSPNGPAGVVSPOPHSSYYSGLPQPPFYNRG 1140
Db 1081 PPLPHGPPRPPAGSYQPPSVCSSTSPNGPAGVVSPOPHSSYYSGLPQPPFYNR 1140
Qy 1141 SGAPGVPVLINLSNVDAVCEKIQIEGLDOSMLPOVCTTIKKNINGRVLQCNIDELK 1200
Db 1141 AVPATGSSLLSSMTVDVCEKIQIEGLDQNMVPOVCTTIKKNINGRVLQCNIDELK 1200
Qy 1201 KEMANFQDWHLPFRSTVLEKRNASHVVPDPRFLSSSSGAPGPHGPARASHNELPHT 1260
Db 1201 KEMANFQDWHLPFRSTVLEKRNASHVVPDPRFLSSSSGAPGPHGPARASHNELPHT 1260

Qy 1261 ELSSQTPYTLNFSFEELNTLGLDEGAPRHSNLSWQSQTRRTTPSLSSLSNQSSSIISKLT 1320
Db 1261 ELSSQTPYTLNFSFEELNTLGLDEGAPRHSNLSWQSQTRRTTPSLSSLSNQSSSIISKLT 1320
Qy 1321 DKVQAEYRDAYREYIAQMSQLEGGPGSTTISGRSPSHSTYYMGSSSGSIHNSLEQKRG 1380
Db 1321 DKVQAEYRDAYREYIAQMSQLEGGPGSTTISGRSPSHSTYYMGSSSGSIHNSLEQKRG 1380
Qy 1381 KDSPEKDDGKHSFLMKRGDVIDYSSSGVSTNDASPLDPITEEDEKSDQSGSKLLPGKKS 1440
Db 1381 KEGELKQEDGKHSFLMKRGDVIDYSSSGVSTNDASPLDPITEEDEKSDQSGSKLLPGKKS 1440
Qy 1441 SERSSLFOTDLKLGSGRLRYOKLSEDESESTESDNTPLKDDKDKAKSGKVERVPKS 1499
Db 1441 SERSSLFOTDLKLGSGRLRYOKLSEDESESTESDNTPLKDDKDKAKSGKVERVPKS 1499
Qy 1500 PEHSAEPIRTPIKAKYLSGLQDPIIARMSICSDSGSVSRSSSPNHSLNHNEVADDSQLEKANLI 1559
Db 1501 QEHSAEPIRTPIKAKYLSGLQDPIIARMSICSDSGSVSRSSSPNHSLNHNEVADDSQLEKANLI 1560
Qy 1560 ELEDSSHGKRGIPHSLSGLQDPIIARMSICSDSGSVSRSSSPNHSLNHNEVADDSQLEKANLI 1619
Db 1561 ELEDSSHGKRGIPHSLSGLQDPIIARMSICSDSGSVSRSSSPNHSLNHNEVADDSQLEKANLI 1620
Qy 1620 LNRTPSTVTLNNSAPANRANQNFDEMEGIRETSQVILRPPSSSPNPTTIONENLKSMTWK 1679
Db 1621 LNRTPSTVTLNNSAPANRANQNFDEMEGIRETSQVILRPPSSSPNPTTIONENLKSMTWK 1680
Qy 1680 RSQRSSYTRLSKDPPELHAAASSESTGFERESIL 1715
Db 1681 RSQRSSYTRLSKDPPELHAAASSESTGFERESIL 1715

RESULT 9

US-10-117-229-3
; Sequence 3, Application US/10117229
; Publication No. US20030190625A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc.
; TITLE OF INVENTION: Human Kidins220PC
; FILE REFERENCE: 9U 104 R1
; CURRENT APPLICATION NUMBER: US/10/117,229
; CURRENT FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 3
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-117-229-3

Query Match 66.4%; Score 5899.5; DB 14; Length 1184;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1138; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 MSVLISSQSVINYYVEENIPALKALLEKCKVDNERNECGQTPLMIAAECQNLIEIVKELIKN 60
Db 1 MSVLISSQSVINYYVEENIPALKALLEKCKVDNERNECGQTPLMIAAECQNLIEIVKELIKN 60
Qy 61 GANCNLEDLONWTALISASKEGHVHIVEELLKCGVNLHHRDMGWTALMWACYKGRDVT 120
Db 61 GANCNLEDLONWTALISASKEGHVHIVEELLKCGVNLHHRDMGWTALMWACYKGRDVT 120
Qy 121 ELLLSHGANSVTVGLQYVYPIIWAAGRGHADIHLLLLONGAKVNCSDKYGTTPVJVAAR 180
Db 121 ELLLSHGANSVTVGLQYVYPIIWAAGRGHADIHLLLLONGAKVNCSDKYGTTPVJVAAR 179
Qy 181 KGHLECVKHLAMGADVQEGANSMTALIIVAKGGYTQSVKEILKRNPNVNLTKDGNNTA 240
Db 181 KGHLECVKHLAMGADVQEGANSMTALIIVAKGGYTQSVKEILKRNPNVNLTKDGNNTA 239
Qy 241 LMIAKSGHTEIVQDLIDAGTVYNIIPDRSGDTVLIIVAGRGHVEIVRALLQYADIDIRG 300

|||||
240 LMIASKEGHEITEIVQDLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRG 299
QY
301 QDNKTALYWAVERKGNATWVRDILQCPDTEICTKDGSETPLIKATKMRNIEVVELLDKGA 360
Db
300 QDNKTALYWAVERKGNATWVRDILQCPDTEICTKDGSETPLIKATKMRNIEVVELLDKGA 359
QY
361 KVSADVKKGDTPLHIAIRGSRKLAELLRNPKDGRLLYRPNKAGETPPYNDICSHOKSIL 420
Db
360 KVSADVKKGDTPLHIAIRGSRKLAELLRNPKDGRLLYRPNKAGETPPYNDICSHOKSIL 419
QY
421 TQIFGARHLSPTETDGMGLGYDLYSSALADILSEPTMQPPICVGLYAQWGSFGSKFLKKL 480
Db
420 TQIFGARHLSPTETDGMGLGYDLYSSALADILSEPTMQPPICVGLYAQWGSFGSKFLKKL 479
QY
481 EDEMKTFAQQIIEPLFQFSLWLVFLTLCCGGLGLLFAFTVHPNLGIAVSLSLFALLYIF 540
Db
480 EDEMKTFAQQIIEPLFQFSLWLVFLTLCCGGLGLLFAFTVHPNLGIAVSLSLFALLYIF 539
QY
541 FIVYFGRREGESWNAWVLSLAEIATLSDACEREFGLATRLFRVFKTEDTQKKKKWKTCCCLPS 600
Db
540 FIVYFGRREGESWNAWVLSLAEIATLSDACEREFGLATRLFRVFKTEDTQKKKKWKTCCCLPS 599
QY
601 TDYNRSSVGGETSLEAMITATLSDACEREFGLATRLFRVFKTEDTQKKKKWKTCCCLPS 660
Db
600 TDYNRSSVGGETSLEAMITATLSDACEREFGLATRLFRVFKTEDTQKKKKWKTCCCLPS 659
QY
661 FVIFLFIIGCIISGITLLAIIRVDPKHLTVNAVLISIASVVGGLAFVLCNCTWQVLDLSL 720
Db
660 FVIFLFIIGCIISGITLLAIIRVDPKHLTVNAVLISIASVVGGLAFVLCNCTWQVLDLSL 719
QY
721 NSQRKELHNAASKHLKSEGFPMVKLCEVELMARMAKTIDSTQNTQRLVWIIDGLDAC 780
Db
720 NSQRKELHNAASKHLKSEGFPMVKLCEVELMARMAKTIDSTQNTQRLVWIIDGLDAC 779
QY
781 EODKVLQMLDTRVRLFSKGPFFIAIFASDPHIIKAINQNLSVLRDSNNGHDMYRNIVH 840
Db
780 EODKVLQMLDTRVRLFSKGPFFIAIFASDPHIIKAINQNLSVLRDSNNGHDMYRNIVH 839
QY
841 LPVFLNSRGLSNARKFLVTSATNGDVPSCDITQIEDADRRVSONSLGEMTKLGSKTALN 900
Db
840 LPVFLNSRGLSNARKFLVTSATNGDVPSCDITQIEDADRRVSONSLGEMTKLGSKTALN 899
QY
901 RRDYRRRQWORTITRQMSFDLTCLKLVTEWDFSDISPDQWRLNIVSVTGRLLRANQIS 960
Db
900 RRDYRRRQWORTITRQMSFDLTCLKLVTEWDFSDISPDQWRLNIVSVTGRLLRANQIS 959
QY
961 FNNDRLASWNLTEQWPPYRSTWILYLEETEGIPDQWTLKTIYERISKNIPTTKDVEPLL 1020
Db
960 FNNDRLASWNLTEQWPPYRSTWILYLEETEGIPDQWTLKTIYERISKNIPTTKDVEPLL 1019
QY
1021 EIDGDIRNFEVFLSSRTPVLVARDVKVFLPCTVNLDPKLEIADVRAAREQISIGLAY 1080
Db
1020 EIDGDIRNFEVFLSSRTPVLVARDVKVFLPCTVNLDPKLEIADVRAAREQISIGLAY 1079
QY
1081 PPLPLHEGPPRAESGYSQPPSCVSTSFNGPPAGGVSPPOPHSSYSGMTGPOHPFYNR 1139
Db
1080 PPLPLHEGPPRAESGYSQPPSCVSTSFNGPPAGGVSPPOPHSSYSGMTGPOHPFYNR 1138

RESULT 10

US-10-117-229-9
; Sequence 9, Application US/10117229
; Publication No. US20030190625A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc.
; TITLE OF INVENTION: Human Kidins220PC
; FILE REFERENCE: 90 104 R1
; CURRENT APPLICATION NUMBER: US/10/117,229
; CURRENT FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9

; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-117-229-9

Query Match 66.4%; Score 5899.5; DB 14; Length 1184;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1138; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 MSVLSIQSVINVEEENIPALKALLEKCKDVBERNECGQTPLMIAAEQGNLEIVKELIKN 60
Db 1 MSVLSIQSVINVEEENIPALKALLEKCKDVBERNECGQTPLMIAAEQGNLEIVKELIKN 60
QY 61 GANCNLEDNNNTALISASKEGHVHIVEELLKCGVNLHRDMGWTALMAWACYKGTDDVV 120
Db 61 GANCNLEDNNNTALISASKEGHVHIVEELLKCGVNLHRDMGWTALMAWACYKGTDDVV 120
QY 121 ELLSHGANPSTVGLQYSVYPIIWAAGRGHADIIVHLLQNGAKVNCSDKYGTTPLVWAAR 180
Db 121 ELLSHGANPSTVGLQYSVYPIIWAAGRGHADIIVHLLQNGAKVNCSDKYGTTPLVWAAR 179
QY 181 KGHLECVKHLAMGADVDOEGANSMTALIVAKGGYTQSVKEILKRNPNVNLTKDGNNTA 240
Db 180 KGHLECVKHLAMGADVDOEGANSMTALIVAKGGYTQSVKEILKRNPNVNLTKDGNNTA 239
QY 241 LMIASKEGHEITEIVQDLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRG 300
Db 240 LMIASKEGHEITEIVQDLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRG 299
QY 301 QDNKTALYWAVERKGNATWVRDILQCPDTEICTKDGSETPLIKATKMRNIEVVELLDKGA 360
Db 300 QDNKTALYWAVERKGNATWVRDILQCPDTEICTKDGSETPLIKATKMRNIEVVELLDKGA 359
QY 361 KVSADVKKGDTPLHIAIRGSRKLAELLRNPKDGRLLYRPNKAGETPPYNDICSHOKSIL 420
Db 360 KVSADVKKGDTPLHIAIRGSRKLAELLRNPKDGRLLYRPNKAGETPPYNDICSHOKSIL 419
QY 421 TQIFGARHLSPTETDGMGLGYDLYSSALADILSEPTMQPPICVGLYAQWGSFGSKFLKKL 480
Db 420 TQIFGARHLSPTETDGMGLGYDLYSSALADILSEPTMQPPICVGLYAQWGSFGSKFLKKL 479
QY 481 EDEMKTFAQQIIEPLFQFSLWLVFLTLCCGGLGLLFAFTVHPNLGIAVSLSLFALLYIF 540
Db 480 EDEMKTFAQQIIEPLFQFSLWLVFLTLCCGGLGLLFAFTVHPNLGIAVSLSLFALLYIF 539
QY 541 FIVYFGRREGESWNAWVLSLAEIATLSDACEREFGLATRLFRVFKTEDTQKKKKWKTCCCLPS 600
Db 540 FIVYFGRREGESWNAWVLSLAEIATLSDACEREFGLATRLFRVFKTEDTQKKKKWKTCCCLPS 599
QY 601 TDYNRSSVGGETSLEAMITATLSDACEREFGLATRLFRVFKTEDTQKKKKWKTCCCLPS 660
Db 600 TDYNRSSVGGETSLEAMITATLSDACEREFGLATRLFRVFKTEDTQKKKKWKTCCCLPS 659
QY 661 FVIFLFIIGCIISGITLLAIIRVDPKHLTVNAVLISIASVVGGLAFVLCNCTWQVLDLSL 720
Db 660 FVIFLFIIGCIISGITLLAIIRVDPKHLTVNAVLISIASVVGGLAFVLCNCTWQVLDLSL 719
QY 721 NSQRKELHNAASKHLKSEGFPMVKLCEVELMARMAKTIDSTQNTQRLVWIIDGLDAC 780
Db 720 NSQRKELHNAASKHLKSEGFPMVKLCEVELMARMAKTIDSTQNTQRLVWIIDGLDAC 779
QY 781 EODKVLQMLDTRVRLFSKGPFFIAIFASDPHIIKAINQNLSVLRDSNNGHDMYRNIVH 840
Db 780 EODKVLQMLDTRVRLFSKGPFFIAIFASDPHIIKAINQNLSVLRDSNNGHDMYRNIVH 839
QY 841 LPVFLNSRGLSNARKFLVTSATNGDVPSCDITQIEDADRRVSONSLGEMTKLGSKTALN 900
Db 840 LPVFLNSRGLSNARKFLVTSATNGDVPSCDITQIEDADRRVSONSLGEMTKLGSKTALN 899
QY 901 RRDYRRRQWORTITRQMSFDLTCLKLVTEWDFSDISPDQWRLNIVSVTGRLLRANQIS 960
Db 900 RRDYRRRQWORTITRQMSFDLTCLKLVTEWDFSDISPDQWRLNIVSVTGRLLRANQIS 959

Qy 961 FNWDLASWNLTEQWPVRTSWLILYLEETEGIPQDMTLKTYIERISKNIPTTKDVEPLL 1020
Db 960 FNWDLASWNLTEQWPVRTSWLILYLEETEGIPQDMTLKTYIERISKNIPTTKDVEPLL 1019
Qy 1021 EIDGDIRNEFVLASRTPLVARDVKVPLPCTVNLDPKRLRIIADVRAAREQISIGGLAY 1080
Db 1020 EIDGDIRNEFVLASRTPLVARDVKVPLPCTVNLDPKRLRIIADVRAAREQISIGGLAY 1079
Qy 1081 PPLPLHEGPRAPSGYSPSPSCSTSPNGFPAGVSPQPHSSYSYSGMTGPQHPFYNR 1139
Db 1080 PPLPLHEGPRAPSGYSPSPSCSTSPNGFPAGVSPQPHSSYSYSGMTGPQHPFYNR 1138
RESULT 11
US-10-106-698-6378
; Sequence 6378, Application US/101066698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6378
; LENGTH: 705
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (244)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (337)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-6378

Query Match 34.3%; Score 3050.5; DB 14; Length 705;
Best Local Similarity 98.8%; Pred. No. 8.9e-192;
Matches 587; Conservative 2; Mismatches 4; Indels 1; Gaps 1;
Qy 62 ANCNLELDNNWTLISASKEGHVHVEELLKCGVNLBHRDGGWTLMWACYKGRTDVVE 121
Db 28 SNCNLELDNNWTLISASKEGHVHVEELLKCGVNLBHRDGGWTLMWACYKGRTDVVE 87
Qy 122 LLLSHGANPSVTGLQYSVYPIIWAAGRCHADIVHLLQNGAKVNCSDKYGTTPPLVWAARK 181
Db 88 LLLSHGANPSVTGL-QYSVYPIIWAAGRCHADIVHLLQNGAKVNCSDKYGTTPPLVWAARK 146
Qy 182 GHLECVKHLAMGADVDEGANSMTALIVAKGVTQSVKBEILKRNPNVNLTDKGNLTAL 241
Db 147 GHLECVKHLAMGADVDEGANSMTALIVAKGVTQSVKBEILKRNPNVNLTDKGNLTAL 206
Qy 242 MIASKEGHEITVDLLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRGQ 301
Db 207 MIASKEGHEITVDLLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRGQ 266
Qy 302 DNKTALYWAKEGNATWVRDILQCNPDTEICTKDGTEPLIKATKORNIIEVVELLDKGA 361
Db 267 DNKTALYWAKEGNATWVRDILQCNPDTEICTKDGTEPLIKATKORNIIEVVELLDKGA 326
Qy 362 VSAVDKKGDTPLHIAIRGRSRKLAELLRNPKDGRLLYRPNKAGETPNIDCSHQSILT 421
Db 327 VSAVDKKGDTPLHIAIRGRSRKLAELLRNPKDGRLLYRPNKAGETPNIDCSHQSILT 386
Qy 422 QIFGARHLSPTETDGMGLYSSALADILSEPTMQPPICVGLYQWGSCKSFLKKLE 481

Db 387 QIFGARHLSPTETDGMGLYSSALADILSEPTMQPPICVGLYQWGSCKSFLKKLE 446
Qy 482 DEMKTEAQOITEPLFQPSWLIIVFLTLCCGGILGFARTVHPNGLIAVLSFLALLYIFF 541
Db 447 DEMKTEAQOITEPLFQPSWLIIVFLTLCCGGILGFARTVHPNGLIAVLSFLALLYIFF 506
Qy 542 IVIYFGRRREGESNNWAWLSTRLARHIGYLELLELLKLMFVNPPELPEOTTKALPVRFLFT 601
Db 507 IVIYFGRRREGESNNWAWLSTRLARHIGYLELLELLKLMFVNPPELPEOTTKALPVRFLFT 566
Qy 602 DYNRLSSVGGTSLAEMIATLSDCEREFGLPFLATRLFRVFKTDTQGGKKKWKKT 655
Db 567 DYNRLSSVGGTSLAEMIATLSDCEREFGLPFLATRLFRVFKTDTQGGKKKWKKT 620

RESULT 12
US-09-835-788A-17
; Sequence 17, Application US/09835788A
; Patent No. US20020077458A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: Death Domain-Containing Receptor Polynucleotides, Polypeptides, a
; FILE REFERENCE: PT018P1
; CURRENT APPLICATION NUMBER: US/09/835,788A
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: PCT/US00/28666
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/159,585
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: 60/167,246
; PRIOR FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-835-788A-17

Query Match 32.1%; Score 2851.5; DB 9; Length 551;
Best Local Similarity 99.6%; Pred. No. 7.5e-179;
Matches 550; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Qy 102 MGGWTALMWACYKGRTDVVELLSHGANPSVTGLQYSVYPIIWAAGRCHADIVHLLQNG 161
Db 1 MGGWTALMWACYKGRTDVVELLSHGANPSVTGL-QYSVYPIIWAAGRCHADIVHLLQNG 59
Qy 162 AKVNCSDKYGTTPPLVWAARKGHLECVKHLAMGADVDEGANSMTALIVAKGVTQSVK 221
Db 60 AKVNCSDKYGTTPPLVWAARKGHLECVKHLAMGADVDEGANSMTALIVAKGVTQSVK 119
Qy 222 EILKRNPNVNLTDKGNLTALMIASKEGHEITVDLLDAGTYVNIIPDRSGDTVLIGAVRG 281
Db 120 EILKRNPNVNLTDKGNLTALMIASKEGHEITVDLLDAGTYVNIIPDRSGDTVLIGAVRG 179
Qy 282 HVEIVRALLQKYADIDIRGQDNKTALYWAKEGNATWVRDILQCNPDTEICTKDGTEPLI 341
Db 180 HVEIVRALLQKYADIDIRGQDNKTALYWAKEGNATWVRDILQCNPDTEICTKDGTEPLI 239
Qy 342 KATKORNIIEVVELLDKGAQVSAVDKKGDTPLHIAIRGRSRKLAELLRNPKDGRLLYRP 401
Db 240 KATKORNIIEVVELLDKGAQVSAVDKKGDTPLHIAIRGRSRKLAELLRNPKDGRLLYRP 299
Qy 402 NKAGETPNIDCSHQSILTQIFGARHLSPTETDGMGLYSSALADILSEPTMQPPI 461
Db 300 NKAGETPNIDCSHQSILTQIFGARHLSPTETDGMGLYSSALADILSEPTMQPPI 359
Qy 462 CVGLYQWGSCKSFLKKLEDEMKTTFAGQOIEPLFQPSWLIIVFLTLCCGGILGFARTV 521
Db 360 CVGLYQWGSCKSFLKKLEDEMKTTFAGQOIEPLFQPSWLIIVFLTLCCGGILGFARTV 419

[illegible]

Search completed: February 5, 2005, 19:54:55
Job time : 147 secs

Job time : 147 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2005, 19:16:39 ; Search time 31 Seconds

(without alignments)
5322.960 Million cell updates/sec

Title: US-10-021-571-4

Perfect score: 8884

Sequence: 1 MSVLISQSVINVEENIPALKECKVDNERECGQTPLMIAAEOGNLEIVKELKN 1715

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5205.5	58.6	1031	2 T43458	hypothetical prote
2	1832.5	20.6	1398	2 T21884	hypothetical prote
3	1255	14.1	246	2 T46392	hypothetical prote
4	555.5	6.3	1881	1 SJHUK	ankyrin 1, erythro
5	552.5	6.2	1856	2 B35049	ankyrin 1, erythro
6	552.5	6.2	1880	2 A35049	ankyrin 1, erythro
7	544	6.1	426	2 AE2149	hypothetical prote
8	539	6.1	3924	2 S37431	ankyrin 2, neurona
9	530	6.0	1961	2 T42716	ankyrin 3, splice
10	529.5	6.0	1940	2 T42715	ankyrin 3, splice
11	525	5.9	1862	2 T49502	ankyrin - mouse
12	523	5.9	1943	2 T42713	ankyrin 3, splice
13	521	5.9	4377	2 A55575	ankyrin 3, long sp
14	516	5.8	1765	2 T42714	ankyrin 3, splice
15	506	5.7	1848	2 S37771	ankyrin, erythrocy
16	475.5	5.4	1549	2 T13940	ankyrin - fruit fl
17	442	5.0	2039	2 T15347	ankyrin-related un
18	415	4.7	2584	2 T24158	hypothetical prote
19	415	4.7	2606	2 T24157	hypothetical prote
20	409.5	4.6	791	2 T42691	hypothetical prote
21	409.5	4.6	1423	1 I37275	death-associated p
22	387.5	4.4	1062	2 T30255	inversin - mouse
23	375.5	4.2	323	2 B47169	ankyrin-like repea
24	373.5	4.2	557	2 T46507	hypothetical prote
25	373.5	4.2	1062	2 T14151	Inv protein - mous
26	349.5	3.9	679	2 T46445	hypothetical prote
27	348	3.9	679	2 A45771	2-5A-dependent RNA
28	343	3.9	741	2 A45771	2-5A-dependent RNA
29	336.5	3.8	991	2 T25412	hypothetical prote

30	329	3.7	1188	2 T19552	hypothetical prote
31	328	3.7	1401	2 S11527	alpha-latrotoxin p
32	325	3.7	934	1 H71274	probable ankyrin
33	320.5	3.6	441	2 S74626	erythroid ankyrin
34	305	3.4	606	2 AC2508	hypothetical prote
35	300.5	3.4	1411	2 S30355	alpha-latroinsecto
36	300.5	3.4	2004	2 D88948	protein ZK1005.1 f
37	300	3.4	1058	2 D82654	ankyrin-like prote
38	293.5	3.3	627	2 E86190	hypothetical prote
39	288.5	3.2	662	2 E84725	ankyrin-like prote
40	288	3.2	1184	2 T00253	gene Ankhzn protei
41	286	3.2	517	2 T48283	ankyrin-like prote
42	285	3.2	368	2 T18184	ankyrin repeat pro
43	285	3.2	633	2 T37499	hypothetical prote
44	284	3.2	1083	2 S48460	probable membrane
45	280	3.2	1001	2 S30385	G9a protein - huma

ALIGNMENTS

RESULT 1

T43458

hypothetical protein DKFZp434F0621.1 - human

C:Species: Homo sapiens (man)

C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004

C:Accession: T43458

submitted to the Protein Sequence Database, December 1999

A:Reference number: Z22517

R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

A:Accession: T43458

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1031 <AAA>

A:Cross-References: UNIPROT:Q9UF42; EMBL:AL133620

A:Experimental source: adult testis; clone DKFZp434F0621

C:Genetics:

A:Note: DKFZp434F0621.1

Query Match 58.6%; Score 5205.5; DB 2; Length 1031;

Best Local Similarity 98.5%; Pred. No. 4.9e-284;

Matches 1008; Conservative 2; Mismatches 0; Indels 13; Gaps 1;

QY	1	MSVLISQSVINVEENIPALKECKVDNERECGQTPLMIAAEOGNLEIVKELKN	60
Db	1	MSVLISQSVINVEENIPALKECKVDNERECGQTPLMIAAEOGNLEIVKELKN	60
QY	61	GANCNLEDLDNWTALISASKEGHVHIVEELKCGVNLHHRDMGGWTALMWACYKGRTDVV	120
Db	61	GANCNLEDLDNWTALISASKEGHVHIVEELKCGVNLHHRDMGGWTALMWACYKGRTDVV	120
QY	121	ELLISHGANPSVTGLQYSVYPIIWAAGRGHADIIVHLLQNGAKVNCSDKYGTTPLVWAAR	180
Db	121	ELLISHGANPSVTGLQYSVYPIIWAAGRGHADIIVHLLQNGAKVNCSDKYGTTPLVWAAR	180
QY	181	KHLCVCKHLLAMGADVDEGANSMTALIIVAKGSGYTOSVKEILKRNPNVNLTKDGNTA	240
Db	181	KHLCVCKHLLAMGADVDEGANSMTALIIVAKGSGYTOSVKEILKRNPNVNLTKDGNTA	240
QY	241	LMIASKEGTEIVQDLLDAGTYVNIIPDRSGDTVLIGAVRGGHVEIVRALLQYADIDIRG	300
Db	241	LMIASKEGTEIVQDLLDAGTYVNIIPDRSGDTVLIGAVRGGHVEIVRALLQYADIDIRG	300
QY	301	QDNKTALYWAVEKGNATVNRDILQCNPTETCTKDGTEPLIKATQNRNIEVELLLDKGA	360
Db	301	QDNKTALYWAVEKGNATVNRDILQCNPTETCTKDGTEPLIKATQNRNIEVELLLDKGA	360
QY	361	KVSADVKKGDPPLHATIGRSKLAELLRNPKGRLLYRNPKAGETPNYDCSHQSKIL	420
Db	361	KVSADVKKGDPPLHATIGRSKLAELLRNPKGRLLYRNPKAGETPNYDCSHQSKIL	420
QY	421	TOIFGARHLSPTETDGMGLDYDLSALADILSEPTMOPPICVGLYAQWGSCKSFLLKKL	480

Qy	1185	NINGRVLAQCNIDELKEMNNNGDWHLFRSTVLEWENAEHVVPDPFRLSSSSGPAP	1244
Db	1120	NLCGLVLATCPLPELKDSMQLPGDWTLIR-LLET-----LKVFGSSP--	1162
Qy	1245	HGEPARRASHNEUPHTLSLSQTPYTLNFSFEELNTLGLDEGAPRHSNLWSQOTRTPSL	1304
Db	1163	-----PGLRVDXRKALTLAEDEEEE-----IEEAEEALN-----SEKERAPLL	1202
Qy	1305	SSL-----NQDSSEISLKTUQVQA-----EYRDAYREVIAQMSOLEGGPGS	1347
Db	1203	GSVRAEQRRRSTIVQNAATELSIDHKCLMEKLSGMDLTETEGDVNEMHFSFSSSTDCP--	1260
Qy	1348	TTISGRSSPHSTYYMGOS-SSGSIH-----SNLEQE-KGKDSPEKDDCKRKFMLKRGD	1400
Db	1261	-----SPWADGFLPASVSAAPSVRFPDDNINDLEASDADSTQSRYSKENVLL-----	1308
Qy	1401	VIDYSSGCVGTNPASPLDPIITEDEKSDQSGKLLPGKSGSSRSLSLFTQDLKLKSGLRY	1460
Db	1309	-----EDERSASPAAHVDMRFDSGNTSSDFVRSYRTS-----SRY	1345
Qy	1461	QKLPDSDESGTBEED	1476
Db	1346	RRMSRGDLMRGMHOND	1361

```

RESULT 3
T46392
hypothetical protein DKFZp434E0620.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C:Accession: T46392
R:Ottenwälder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23031
A:Accession: T46392
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-246 <NAA>
A:cross-references: UNIPROT:Q9NT37; EMBL:AL137553
A:Experimental source: adult testis; clone DKFZp434E0620
C:Genetics:
A>Note: DKFZp434E0620.1

```

Query Match	14.1%;	Score 1255;	DB 2;	Length 246;
Best Local Similarity	99.6%;	Pred. No. 3e-63;		
Matches 246;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	1470	SGTEESDNTPLLKDDKDKRKAEGKVERVPKSPHSAEPIRTFIKAEVLSDALLDKXSSD	1529	
Db				
Db	1	SGTEESDNTPLLKDDKDKRKAEGKVERVPKSPHSAEPIRTFIKAEVLSDALLDKXSSD	60	
Qy	1530	SGVRSSESSPNHSLJHNEVADDSOLEKANLIELEDHDSHGSKRGIPHSISGQDDPIIARMSI	1589	
Db				
Db	61	SGVRSSESSPNHSLJHNEVADDSHLEKANLIELEDHDSHGSKRGIPHSISGQDDPIIARMSI	120	
Qy	1590	CSDKIKSPSCSLTIASSPEENWPACQKAYNLNRTPTSVTLNNNSAPANRANQNFDEMEGI	1649	
Db				
Db	121	CSEDKIKSPSCSLTIASSPEENWPACQKAYNLNRTPTSVTLNNNSAPANRANQNFDEMEGI	180	
Qy	1650	RETSQVILRPSSSPNPTTIQENLKGMTKXRSQRSSYTRLSDKPPPELHAAASSESTGFGGE	1709	
Db				
Db	181	RETSQVILRPSSSPNPTTIQENLKGMTKXRSQRSSYTRLSDKPPPELHAAASSESTGFGGE	240	
Qy	1710	ERESIL 1715		
Db				
Db	241	ERESIL 246		

RESULT 4
SJHUK
ankyrin 1, erythrocyte splice form 1 - human
N;Alternate names: ankyrin 2.1, erythrocyte; ankyrin-R

N;Contains: ankyrin 2.2
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text_change 09-Jul-2004
C;Accession: S08275; A33219; PC2220; A35443
R;Lux, S.E.; John, K.M.; Bennett, V.
Nature 344, 36-42, 1990
A;Title: Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure w
A;Reference number: S08275; MUID:90158830; PMID:2137557
A;Accession: S08275
A;Molecule type: mRNA
A;Residues: 1-1881 <LU1>
A;Cross-references: UNIPROT:P16157; EMBL:X16609; NID:g28701; PIDN:CAA34610.1; PID:g28702
A;Accession: A33219
A;Molecule type: protein
A;Residues: 2-7, 'X', 9-17, 'X', 19-20, 'T', 22-30, 733-749, 'A', 751-753, 828-833, 'X', 835-855, 'X',
X', 1367, 1383-1427, 1601-1630, 1686-1698, 'D', 1700; 1763-1772 <LUX>
A;Note: 845-Arg and 1392-Thr were also found
R;Hermann, J.; Barel, M.; Frade, R.
Biochem. Biophys. Res. Commun. 204, 453-460, 1994
A;Title: Human erythrocyte ankyrin, a cytoskeleton component, generates the p57 membrane
A;Reference number: PC2220; MUID:95071348; PMID:7526850
A;Accession: PC2220
A;Molecule type: protein
A;Residues: 910-929 <HER>
R;Davis, L.H.; Bennett, V.
J. Biol. Chem. 265, 10589-10596, 1990
A;Title: Mapping the binding sites of human erythrocyte ankyrin for the anion exchanger
A;Reference number: A35443; MUID:90285190; PMID:2141335
A;Accession: A35443
A;Molecule type: protein
A;Residues: 'X', 5, 'X', 7-12, 403-417, 'X', 419-422, 'H', 424, 'LQ', 797-800, 'L', 802-814; 862-863,
C;Genetics:
A;Gene: GDB:ANK1; ANK
A;Cross-references: GDB:118737; OMIM:182900
A;Map position: 9p11.2-9p11.2
C;Superfamily: ankyrin; ankyrin repeat homology
C;Keywords: alternative splicing; phosphoprotein
F;2-1881/Product: ankyrin 1, erythrocyte form 1 #status predicted <MAT1>
F;2-1532,1675-1881/Product: ankyrin 2.2, erythrocyte #status predicted <MAT2>
F;2-827/Domain: 89K #status predicted <DOM1>
F;2-827/Region: anion exchange protein binding
F;44-76/Domain: ankyrin repeat homology <AN01>
F;77-109/Domain: ankyrin repeat homology <AN02>
F;110-142/Domain: ankyrin repeat homology <AN03>
F;143-171/Domain: ankyrin repeat homology <AN04>
F;172-204/Domain: ankyrin repeat homology <AN05>
F;205-237/Domain: ankyrin repeat homology <AN06>
F;238-270/Domain: ankyrin repeat homology <AN07>
F;271-303/Domain: ankyrin repeat homology <AN08>
F;304-336/Domain: ankyrin repeat homology <AN09>
F;337-369/Domain: ankyrin repeat homology <AN10>
F;370-402/Domain: ankyrin repeat homology <AN11>
F;403-435/Domain: ankyrin repeat homology <AN12>
F;436-468/Domain: ankyrin repeat homology <AN13>
F;469-501/Domain: ankyrin repeat homology <AN14>
F;502-534/Domain: ankyrin repeat homology <AN15>
F;535-567/Domain: ankyrin repeat homology <AN16>
F;568-600/Domain: ankyrin repeat homology <AN17>
F;601-633/Domain: ankyrin repeat homology <AN18>
F;634-666/Domain: ankyrin repeat homology <AN19>
F;667-699/Domain: ankyrin repeat homology <AN20>
F;700-732/Domain: ankyrin repeat homology <AN21>
F;733-765/Domain: ankyrin repeat homology <AN22>
F;766-798/Domain: ankyrin repeat homology <AN23>
F;828-1382/Domain: 62K #status predicted <DOM2>
F;1383-1881/Domain: 55K #status predicted <DOM3>
F;828-1382/Region: spectrin binding

Query Match 6.3%; Score 555.5; DB 1; Length 1881;
Best Local Similarity 19.7%; Pred. No. 1.6e-22;
Matches 374; Conservative 286; Mismatches 659; Indels 581; Gaps 72;
QY 21 LKALLEKCKVDVRNCGGOTPLMAAQQNKLETIVKELIKNGANCNULEDNWTALISASK 80

QY	519	FTVHNELGIAVSLSEFALLXYFFIVIFGGRREGSWMNAWVLSTRLARHIGYLELLK-	577
Db	657	KQANGNLGNKSGLTFLHLV-----AQEG-----HVPADVLIKH	690
QY	578	-----LMFNVPPELPEOTTKALPVRFPLP---TDYNRLSSVG-----GETSLA	616
Db	691	GUMVDATTMRGYTPPLHVASHYGNIKLVKLFLLQHQAADVNAKTGLGYSPLHQAAQQOHTDIV	750
QY	617	EMI-----ATLSDACEREFGLA--TRLFRPKTEDTGKKWKKTCCLPSPVF	662
Db	751	TLLLKNAGASPNEVSSDGTTPILAIAKRGLIISVTDLKVU-TDEI-----SVF	796
QY	663	IPLFIIGICIISGITULLAIERYDPKHL-----TVNAVLSIASVVG---LAFVLCNRTMWQ	714
Db	797	L-----VSDKRMSPPETVDEIL-DVSEGEELISFAERR-----832	
QY	715	VLDSSLNSQRKLHNAASKLHKLKSEGPMVKLKEVELMARMAKTIIDFTQNQLRVUVII	774
Db	833	--DSRDVDEEKELLDFVKLDQVVESPAIRIPC-----AMPETVVISRSEQEASKEY	884
QY	775	DGLDACEQDKVLOMLDTVRVLFS--KGPFIAIFASDP-----HIHK-----814	
Db	895	DEDSLIPSSPATETSDNISPVASPVHTGFLVSFMVDARGGSMRGSRHNLRVVIPPTCA	944
QY	815	-----AINONLSNVLRDSNINGHDYMRN-IVHLVPFL-NSRGLS	851
Db	945	APTRITCELVKPKLSTPPPLAEBEGLASRIIALGPTGAOFLSPVIVEIPHASHCRG--	1000
QY	852	NARKPLVTSANGDVPCSDTTGIOEDADRRVSQNSLGEMTKLGSKALTNRDRTYRRRQM	911
Db	1003	-DRELAVLRSNGSV-----WKEHRSTRYESYLDOILN-GMDEELGSLEEKKRVC	1050
QY	912	RTITRQMSFDLTCLLVAT---EDWESDISPOTMRLLNIVSVTGRLLIRANOQISFNWDLRAS	968
Db	1053	RILT--TDFPLYFWIMSLCDD-YDIIGPEG-----GSLKSLVPLVQVATFPENAVTK	1100
QY	969	WINTEQWPYRTSWLLILETEGIPDOMTLK-----TIVERISK-----1000	
Db	1103	RVKLAQO-----AQVPDELVTCLKGNQATFPSIVTEPRRRKFHPRPIGLR	1144
QY	1009	-NIPTTKDVPELLEIDGINFEVF-----LSRTPVLVARDVKVFLPCTV	1050
Db	1149	IPLPSSWTNDRDGEGDTTSLRLSCSVIGGTDAQWEDITGTTKLVYANECAF--TT	1200
QY	1054	NLDPKLREIADVRAAREQISIGLAYPPLPHHEGPAPPAGSYQSPPSVCSTSFNGPFA	1111
Db	1206	NV--SARFWLSDCPRTAEAVNFATLLYKEL-----TAPVMKXVF	1244
QY	1114	GGVSPQPHSSYYSGMTQPQHPFYNRGSGPAPGV--VLLNSLNVDACE-----K	1160
Db	1244	-----IFAKNDPREGLRCYCMTDDKDVKTLQEHENVEVAR	1280
QY	1163	LKOIEGLDQSMLOYCTTIKANINGRVLACNQIDELAKENMNMFWDHFLFRSTVLEMNR	1222
Db	1282	SRDIEVLEGMSL-----PAELSG-----NLVPVKKAQORSFHQSFRENLAM--	1322
QY	1223	AESHVVPEDRFLSESSGGPAPHCEPARASHNE-----LPHTELS-----1260	
Db	1326	-----PVKVRDSSREPGGSLSFTRKAMKYEDTOHILCHLNIITMPPCAAGSAEDRR	1370
QY	1264	-SQPYTLNFSFEELNTLGLDEGAPR-----HSNLWSQSTRTP-SLSSLSNQD	1311
Db	1377	RTPLPALRYLSILSESTPGSLUGTEQEAEMKNAVISEHLGSLWAELARLQFSVEDIN---	1430
QY	1312	SSBISKLTDKVAERYDAYREYIAQMSQL-----EGGPGSTTISRSSPSHYTMQGSSS	1360
Db	1434	-----RIRVENPNLSLLEQSVALLNLWVIREG-----QNAVNNLYTALQSID	1475
QY	1368	GGSIHSHNLEQBKGKDBPKPDGRKRSFKMKRGDVIDYSSSGVSVTNDAASPLPOITEEDS-	1425
Db	1476	RGEIVNMLEGSGQRSNRLKPD-----RRHTDRDYSLSPQMGYSYL-----QDBLL	1522

QY 1426 -----KSDQSGKL-----LPGKSSRSLSLFTQDLKLKSGSLRQKPLPSDE 1467
Db 1523 SPASLGALSSPLRADQYNEVAILEDALPLAATHEHTMLEMSDMQVWSAGLTPLSLVTAED 1582
QY 1468 DE---SCTRESNTPLLKODKDKABCKVERVPKSPHSAEPIRTFIKAKEYLSLALLOK 1524
Db 1583 SSLECSKAEDSDAT-----GHEWKEGALSEERPRPE-----LGSILEVEDDTVDS 1628
QY 1525 KQSS-----DSGVRSSSSNNHLSHNEVADDSQLEKANLIELEDDSHSGKRGKPHSL 1576
Db 1629 DATNGLIDLEQEGORSEKPLGSKRQDDATGAGQ-DSENEVSLVSGHQGQARITHS- 1686
QY 1577 SGLQDPIIARMSICSEDK-----KSPSECSLIASSPEENW 1611
Db 1687 -----PTVSQVTRSQDRLOQDWADGSIYSYLODAAQGSW 1721

RESULT 6

A35049
N;Alternate names: ankyrin 2.1, erythrocyte; ankyrin-R
N;Contains: ankyrin 2.2, erythrocyte
C;Species: Homo sapiens (man)
C;Date: 27-Jul-1990 #sequence_revision 01-Oct-1992 #text_change 09-Jul-2004
C;Accession: A35049
R;Lambert, S.; Yu, H.; Prchal, J.T.; Lawler, J.; Ruff, P.; Speicher, M.C.; K
Proc. Natl. Acad. Sci. U.S.A. 87, 1730-1734, 1990
A;Title: cDNA sequence for human erythrocyte ankyrin.
A;Reference number: A35049; MUID:90175370; PMID:1689849
A;Accession: A35049
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1880 <LAMB>
A;Cross-references: UNIPROT:P16157; GB:M28880
C;Genetics:
A;Gene: GDB:ANK1; ANK
A;Cross-references: GDB:118737; OMIM:182900
A;Map position: 8p11.2-8p11.2
C;Superfamily: ankyrin; ankyrin repeat homology
C;Keywords: alternative splicing; cytoskeleton
F;2-1880/Product: ankyrin 1, erythrocyte form 2 #status predicted <MAT>
F;2-1513,1676-1880/Product: ankyrin 2.2, erythrocyte #status predicted <MA2>
F;44-76/Domain: ankyrin repeat homology <AN01>
F;77-109/Domain: ankyrin repeat homology <AN02>
F;110-142/Domain: ankyrin repeat homology <AN03>
F;143-171/Domain: ankyrin repeat homology <AN04>
F;172-204/Domain: ankyrin repeat homology <AN05>
F;205-237/Domain: ankyrin repeat homology <AN06>
F;238-270/Domain: ankyrin repeat homology <AN07>
F;271-303/Domain: ankyrin repeat homology <AN08>
F;304-336/Domain: ankyrin repeat homology <AN09>
F;337-369/Domain: ankyrin repeat homology <AN10>
F;370-402/Domain: ankyrin repeat homology <AN11>
F;403-435/Domain: ankyrin repeat homology <AN12>
F;436-468/Domain: ankyrin repeat homology <AN13>
F;469-501/Domain: ankyrin repeat homology <AN14>
F;502-534/Domain: ankyrin repeat homology <AN15>
F;535-567/Domain: ankyrin repeat homology <AN16>
F;568-600/Domain: ankyrin repeat homology <AN17>
F;601-633/Domain: ankyrin repeat homology <AN18>
F;634-666/Domain: ankyrin repeat homology <AN19>
F;667-699/Domain: ankyrin repeat homology <AN20>
F;700-732/Domain: ankyrin repeat homology <AN21>
F;733-765/Domain: ankyrin repeat homology <AN22>
F;766-798/Domain: ankyrin repeat homology <AN23>

Query Match 6.2%; Score 552.5; DB 2; Length 1880;

Best Local Similarity 19.6%; Pred. No. 2.4e-22;

Matches 373; Conservative 287; Mismatches 659; Indels 581; Gaps 72;

QY 21 LKALBKCKDVBNERCGQTFPLMAAQGNLEIVKELIKNGANCNLEDLDNWTALISASK 80

Db 94 VRELVNYGANVNAQSKGFTPLYMAAQENHLEVVKFLENGANQNVATEDGFTPLAVALQ 153

QY 81 EGHVHIVEELLKGV-----NLEHRDMCGWTALMWA 111
Db 154 QGHENVAHLINYGTKGVRLPALHIAARNDDTRTAAVLLQNDPNDVLSKGTGFTPLHIA 213
QY 112 CYKGRTDVVELLSHGANPSVT-----GLQY-----SV 139
Db 214 AHYENLNVAQLLNRGSSVNFPPQNGITPLHIASRRGNVIMVRLLLDRGAQIETKTKDEL 273
QY 140 YPIIWAAGHADIIVHLLONGAKVNCSDKYGTTPLVWAARKGHLECEVKKHLLAMGADVD- 198
Db 274 TPLHCAARNGHVRISILLDHGAPIQAKTKNGLSPIHMAAQGDHLDVRLQLLYQDAEIDD 333
QY 199 -----QEGA-----NSMTALIVAKGYTQSVKILKE 226
Db 334 ITLDHLTPLHVAACHGHHRVAKVLLDKGAKPNRSLNGFTPLHIACKKHVRVWELLKT 393
QY 227 NPNVNLTDKQNTALMIASKEGHEITIVODLLDAGTYVNIIPDRSGDTVLICAVRGGHVEIV 286
Db 394 GASIDAVTESGLTPLHVASFMGHLPIVKNLLQRGASPNVSNVKVETPLHMAABAGHTEVA 453
QY 287 RALLOKYADIDIRGQDKNTALYWAKEGNATWVRDILQCNPDTEICTKOGETPLIATKM 346
Db 454 KYLLQNKAKVNAKXODQTPHCAARIGHTNMVKLLLENNANPNLATAGHTPLHIAARE 513
QY 347 RNIEVVELLDKGAHSVADYKKGDTPLHIAIRCSRKLAELLLRNPKDGRLLVYPNKAGE 406
Db 514 GHVETVLLALLEKEASQACMTKKGFTPLHVAAKYKVRVAELLELLE--RDA-----HPNAAGK 567
QY 407 ---TPYNIDCSHQKSILTQIFGARHLSPTETDGDMLGY-DLYSSALADILSEPTMOPPIC 462
Db 568 NGITPLHVAHNNLDIVKLLPRGSGSPHSPAWN--GYTPLHIAAKQN-----QVEVA 618
QY 463 VGLYAQWGGSGFLKLEDEMKTFAGQOIEPLFQFSWLIIVFLTLILCGG-----LGILFA 518
Db 619 RSLLOYGGSANAESV-----QGVTPHL-----LAAQEGHAEMVALLS 656
QY 519 FTVHNHGLTAVSLFLALYIFVIVYFGRREGESNNWAVLSTRLAHIGVLELLK- 577
Db 657 KQANGNLGNKSGLTPLHLV-----AQEG-----HVPVADVLIIKH 690
QY 578 -----LMFVNPPELPQTTKALPVRFLF---TDYNRLSSVVG-----GETSLA 616
Db 691 GVMVDATTENGTYPLHVASHYGNIKLVKLLQHQADVNAKTGLGYSPLHQAQOQHTDIV 750
QY 617 EMI-----ATLSDACEREFGLA-TRLFRVPKTEDTQKKKWKTCCLPSPFV 662
Db 751 TLILLKNGASPNESVSDGTTPLAIAKRLGYISVTDVLKV-VTDET-----SFV 796
QY 663 IFLFIIGCIISGITLLAIFRVDPKHL-----TVNAVLISIASVVG---LAFVLNCRTWQ 714
Db 797 L-----VSDKHMSPPETVDEIL-DVSEDEGELISFKAER- 832
QY 715 VLDSLINSQRKLIHNAASKLHLKSEGFMKVLKCEVELMARMAKTIDSTQNTQRLVVI 774
Db 833 --DSRDVDEKELLDFVKLDQVVESPAIPRIPC-----AMPETVIRSEEQEQASKEY 884
QY 775 DGLDACEQDKVQLMDTVRVLFS--KGPFIATIPASDP-----HIILK-----814
Db 885 DEDSLIPSSPATETSDNISPVASPVHTGFLVSPMDARGSMGRSHRNLGRVVPRTCA 944
QY 815 -----AINQNLNSVLRDSNINHDIYMN-IVHLPVEL-NSRGLS 851
Db 945 APTRTICRLVPOKUSTPTPLAEEBGLASRIIALGPTGAQFLSPVIVEIPHAFASHG-- 1002
QY 852 NARKFLVTSATNGDVPSCSDTTGQEDADRRVSONSLGEMTKLGSKTALNRDRYRRQMQ 911
Db 1003 -DRELVLASENGSV-----WKEHRSRYGESYLDQILN-GWDEELGSLSELEKRVK 1052
QY 912 RTITRQMSFDLTKLIVT---EDWFSDISQTMRRLLINIVSVTGRLLRANQISNWRDLAS 968
Db 1053 RIIT--TDPFLVFMVMSRLCQD--YDIIGPEG-----GSLKSKLVPLVQATPENAVTK 1102

Qy	969	WINLTEQWPRYTSMLIILYLBTEGIPQMVLK-----TIYERISK-----	1008
Db	1103	RVKLALQ-----AQVPVDELTKLGNOATFSPIVTBPRRKFHRPIGLR	1148
Qy	1009	NIPTTKDVEPLLEIDGDIRNFEP-----LSSRTFVLAVARDKVFLPCTV	1053
Db	1149	IPLPPSWTDNPRDSGEGETTSLRLLCVSIGGTDOAQWEDITGITTKLVANECAF--TT	1205
Qy	1054	NLDPKLREIIADVRAAREQISIGLAYPPLPLEHGPAPRASGYSPQSVCSTSFNGPFA	1113
Db	1206	NV--SARFWLSDCPRTAEAVNFATLLYKEL-----TAVPYMAKFV	1243
Qy	1114	GGVVSPQHSHSYSGMTGPQHPFYNRGSGPAGPV--VLLNSLNVDACE-----X	1162
Db	1244	-----IFAKMNDPREGRLCYCMTDDKXTLEQHENFEVAR	1281
Qy	1163	LKOIEGLDQSMLPQYCTTIKKANINGRVLAQCNIDELKEMNMNFGDWHLFRSTVLEMNR	1222
Db	1282	SRDIEVLEGMSL-----PAELSG--NLVPVKAAQQRSFHQSFRENRLAM--	1325
Qy	1223	AESHVVPEDPRFIJSESSSGPHAPGEPARRASHNE-----LPHTELS-----	1263
Db	1326	-----PVKVRDSSREFPGSLSPLRKAMKYEDTOHILCHLNITMPPCAKGGAEDRR	1376
Qy	1264	-SQTPYTLNFSFEELNTLGLDBGNR-----HSNLSWSQOTRPT-SULSNSQD	1311
Db	1377	RTPTPLALRYSIJSESTPGSLSGTEQAEMRWAVISHEHLGSLWAELARELOFSVEDIN---	1433
Qy	1312	SSIETSKLTOKVAEYRDAYREYIAOMSOL---EGPGGTISIGRSSPHSTYYMGQSSS	1367
Db	1434	-----KIRVENPSLLEGSVALNLNWIREG-----ONANMENLTAQSID	1475
Qy	1368	GGSIHNSLEOEKGKSDPEPKDCKRPFMKRGDVIDSYSVGSVTNDAASPLDPITEDE--	1425
Db	1476	RGEIVNMLEGSQRGRNLKPD-----RHHTDRDYLSFPSQMWGVSSL-----QDELL	1522
Qy	1426	-----KSDQGSKL-----LP GKKSERSSSLFQTDLKLGSGLRKYQLPSDE	1467
Db	1523	SPASLGCALSPLRADQYWNVEAILDAIPLAATEHDTMLEMSDMQVMWSAGLTPLSLVTAED	1582
Qy	1468	DE---SGTESSDNTPLLKDKRKAEGKVERVPKSPHESAERTPTFKAEYLSDALLDK	1524
Db	1583	SSLSCSAEDSDAT-----GHEWKLEGALSEEPRGPB-----LGSJLELVEDTVD	1628
Qy	1525	KDSS-----DSGVRSSSESNPHLSHNVEADDSOLEKANLIJELEDSDSHSGKGI PHSL	1576
Db	1629	DATNGLIDLLEQEBQGRSEBKPGSKKQDDATGAGO-DSENEVSLVSGHQRGQARITHS-	1686
Qy	1577	SLGDPTIARMSCISEDK-----KSPSECSLIASPENN	1611
Db	1687	-----PTVSOVTERSODRLQWDADGSGIVSLOYDAAOGSW	1721

RESULT 7

AE21149
hypothetical protein all12748 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Name: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AE2149
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AE2149
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-426 <KUR>
A;Cross-references: UNIPROT:Q8VTG9; GB:BA0000019; PIDN:BAB74447.1; PID:gl7131841; GSPDB:C
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all2748

Query Match	6.1%	Score 544;	DB 2;	Length 426;
Best Local Similarity	37.2%;	Pred. No. 6.2e-23;		
Matches 149;	Conservative 64;	Mismatches 180;	Indels 8;	Gaps 6
Qy	9	VINVEEENIPALKALLEKCKDVERNECGOTPLMIAAEOGNLHVIVKELIKNGANCNL-E	67	
Db	8	LLKVAKSGDIKGLGALLAAGGVVDICRBDGTTALMFAANLGYTHIVRSLLDGGANVNLAR	67	
Qy	68	DLDNWTALISASKEGHVHIVVELLKCGVNLHRDMGWTALMWACYKGRDTDVVELLSHG	127	
Db	68	KRYGLTALMLAASANQVDIVQLLISRGAAVNATNEDGSTALMAAALKGNVEVARVLLAAG	127	
Qy	128	ANPSVTGLQISVYPIIWAAGRGHADIVHLLLONGAKVNCSPKYCTTPLVWAARKHLECV	187	
Db	128	ADVNTDKDDT-ALKLAVKRGQAQVQLILPUSDADANCEDEEGETLLMLAASGHGDVV	186	
Qy	188	KHLLAMGADVQEGANSMTALIVAVKGGYTQTSVKEILKRNPNVNLTKDGNLTALMIASKE	247	
Db	187	QVLLAAGVDVNEQNDGSTALLAAVAAGNAIKILLDRGADVNHQDQGESALHLATVE	246	
Qy	248	GHTHEIVODLLDAGTYVNIPIRBSGDTVLIGAVRGHVEIVRALLQYADIDITRGD-NKTA	306	
Db	247	GYVDVQVLLNQAGANTQIKNKLGBTPLLVAAALQHQDQIVETLL-KYG-ANVHGDNLGETP	304	
Qy	307	LYXWAVEKGNATVWRDIIQCNPDTETICTKDGETPLIKATKRNIEWELLDDKGAKVSAVD	366	
Db	305	LTLAASQGHATVRIILDDYGNANI PASDGTALIKATERNHPGVIQULLAKGANVNVQD	364	
Qy	367	KKGDTPLHIAIRGRSRKLABLLRNPKDGRLLLYRPNKAGET	407	
Db	365	SVGATALIIWAASGGYNKVVQVLLIEGGADTNL---KNRGGYT	402	

RESULT 8

S37431
N:Ankyrin 2, neuronal long splice form - human
N:Alternate names: ankyrin B, 400K splice form; ankyrin-B; brain ankyrin; non-erythroid
N:Contains: ankyrin 2, short form
C:Species: Homo sapiens (man)
C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 09-Jul-2004
C:Accession: S37431; A39643; B39643; A40334; A49462; S14533; S14569
R:Chan, W.
submitted to the EMBL Data Library, September 1993
A:Reference number: S37431
A:Accession: S37431
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3924 <CHA>
A:Cross-references: UNIPROT:Q01484; EMBL:Z26634; NID:g406287; PIDN:CAA81387.1; PID:g406306
R:Otto, E.; Kunimoto, M.; McLaughlin, T.; Bennett, V.
J. Cell Biol. 114, 241-253, 1991
A:Title: Isolation and characterization of cDNAs encoding human brain ankyrins reveal a
A:Reference number: A39643; MUID:91302466; PMID:1830053
A:Accession: A39643
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2077 <OTL>
A:Cross-references: GB:X56957
A:Accession: B39643
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1443,3585-3924 <OTT>
A:Cross-references: EMBL:X36958
R:Tse, W.T.; Menninger, J.C.; Yang-Feng, T.L.; Francke, U.; Sahr, K.E.; Lux, S.E.; Ward,
Genomics 10, 858-866, 1991
A:Title: Isolation and chromosomal localization of a novel nonerythroid ankyrin gene.
A:Reference number: A40334; MUID:92009921; PMID:1833308
A:Accession: A40334
A:Molecule type: DNA
A:Residues: 463-474, PE', 477-495 <TSE>
A:Cross-references: GB:M37123; NID:g178647; PIDN:AAA62828.1; PID:g178648
R:Chan, W.; Kordeli, E.; Bennett, V.

J. Cell Biol. 123, 1463-1473, 1993

A>Title: 440-kD ankryrinB: structure of the major developmentally regulated domain and sequence

A/Reference number: A49462; MUID:94075409; PMID:8253844

A/Accession: A49462

A/Status: preliminary; nucleic acid sequence not shown

A/Molecule type: mRNA

A/Residues: 1-3924 <RES>

A/Cross-references: EMBL:Z26634; NID:g406287; PIDN:CAA81387.1; PID:g406288

C/Genetics:

A/Gene: GDB:ANK2

A/Cross-references: GDB:127607; OMIM:106410

A/Map position: 4q25-4q27

C/Superfamily: ankryrin; ankryrin repeat homology

C/Keywords: alternative splicing

F/2-3924/Product: ankryrin 2, long form #status predicted <MAT>

F/2-1443, 3585-3924/Product: ankryrin 2, short form #status predicted <MA2>

F/63-95/Domain: ankryrin repeat homology <AN01>

F/96-128/Domain: ankryrin repeat homology <AN02>

F/129-161/Domain: ankryrin repeat homology <AN03>

F/162-190/Domain: ankryrin repeat homology <AN04>

F/191-223/Domain: ankryrin repeat homology <AN05>

F/232-264/Domain: ankryrin repeat homology <AN06>

F/265-297/Domain: ankryrin repeat homology <AN07>

F/298-330/Domain: ankryrin repeat homology <AN08>

F/331-363/Domain: ankryrin repeat homology <AN09>

F/364-396/Domain: ankryrin repeat homology <AN10>

F/397-429/Domain: ankryrin repeat homology <AN11>

F/430-462/Domain: ankryrin repeat homology <AN12>

F/463-495/Domain: ankryrin repeat homology <AN13>

F/496-528/Domain: ankryrin repeat homology <AN14>

F/529-561/Domain: ankryrin repeat homology <AN15>

F/562-594/Domain: ankryrin repeat homology <AN16>

F/595-627/Domain: ankryrin repeat homology <AN17>

F/628-660/Domain: ankryrin repeat homology <AN18>

F/661-693/Domain: ankryrin repeat homology <AN19>

F/694-726/Domain: ankryrin repeat homology <AN20>

F/727-759/Domain: ankryrin repeat homology <AN21>

F/760-792/Domain: ankryrin repeat homology <AN22>

F/793-825/Domain: ankryrin repeat homology <AN23>

Query Match	6.1%;	Score 539;	DB 2;	Length 3924;
Best Local Similarity	20.1%;	Pred. No. 4.6e-21;		
Matches	421;	Conservative 294;	Mismatches 717;	Indels 662; Gaps 81;

Qy	14	BEENIPALKALLEKCKVDNERNEGQOTPLMTAAEOGNLEIVKELIKNGANCNLELDNDWT	73
Db	73	KEGHVGLVOELLGRGSSVDSATKGNLTALHIASLAGOAEVVKVLKEGANINASONGFT	132
Qy	74	ALISASKEGHVHIVVEELKCGVN	103
Db	133	PLYMAAQENHIDVVVKYLLENGANGQSTATEDGFTPLAVALQOQHNOQVAILLENDTKGVR	192
Qy	104		123
Db	193	LPALHIAARKDDTSAALLQNDHNADVQSRKMVNRTTESGFTPLHIAAHYGNVNVAFL	252
Qy	124	LSHGANSVTLGQTSVYPIIWAAGRGHADI VHLHLIQNGAKYNCSDKYGTTPLVWAARKG	182
Db	253	LNRRGAADVFTA-RNGITPLHVASKRGNTNMVKLLLDRGQIDAKTRDGLTPLHCAARSGH	311
Qy	183		210
Db	312	DQVVELLERLARGAPLARTKNGLSPLHMAAQGDHVECVKHLHQHKAFPDVTDLVDYLTAHV	371
Qy	211	AVKGGYQTSVKEILKRNPNVNLTKDGNLTALMIASKE	247
Db	372	AAHCGHYRVTKLLDKRANPNARALNGFTPLHIAACKNRIKVMELLVKYGAISOAITESG	431
Qy	248		297
Db	432	LTPIHVAAFWMGHUINVLLLLQNGASPDVTNIRGETALHMAARAGQVVRCLLRNGALVD	491
Qy	298	IRGQDNKTALYWAWEKGNATVWRDILQCNPOTEICTDKGETPLIKATKORNIEVVVELLD	357

Db	492	ARAREEQTPLHIA	SRGKTEIVQLLQ	MAHPDAA	TNGYTP	PLHIGAREQ	QVDVASV	LE 551														
Qy	358	KGAKVSAVDK	GGDTPPLHIA	IRGRSRK	LAELLLR	---	---	NPXD 394														
Db	552	AGAAHSLATK	GGTTPPLHIA	KYSGSLD	VAKLLL	ORRAA	ADSAGKNGITPLHVA	AHYDNQKV 611														
Qy	395	GRLL	---	YRPNKAGET	FPYNIDC	SHQKSILTQI	--FGAR	-----HLSPT 432														
Db	612	ALLLLEK	GASPHATAK	NGYTPPLHIA	AKONQ	MAJSTLLN	YGAE	TNIVTKQGVTPPLH	LASQ 671													
Qy	433	ETGDDML	---	GYDLYSSALADI	SEP	TMQPPI	CVGLYAQWGS	KSFLLK	LEDEM-- 484													
Db	672	EGHTDVTLL	DLKXANI	HM	TKSGLTS	---	---	---	---	EDKVVN 710												
Qy	485	---	---	KTFGAQ	QIEPLFPQ	SM	LIVFLTLLC	--GGLGLL	FAFTVHPNL	GIASVLSFLAL 536												
Db	711	ADILTRHGA	ODDAHTK	GLGYTP	PLIV	---	---	ACHYGNV	KMVNFLKQ	GANVNAKTYNGT	TPPL 765											
Qy	537	---	---	LYIFFTVI	YFGREGE	SNMAW	LSVTLRA	HIGYLELL	--LKL	----- 579												
Db	766	HQAAQOQ	GHTHINV	LQHGAK	PNATTANG	TALA	--IAKRLG	IVISVVD	LTKV	TEEVTTT 823												
Qy	580	---	---	---	---	---	---	FVNPP	ELPQTTKALP	--VRPL-F 600												
Db	824	TTTTITEKH	KLNPET	WTEVL	DSDEB	DDTMTG	DGGEYLR	PEDLKEL	GDLS	SPSQFLDG 883												
Qy	601	TDYNRLSS	VGGET	SL	---	---	---	---	---	---	AEMLATLS	DACERE 629										
Db	884	MNYLRYS	LEGBR	SDSR	SPSSDR	SH	TLASHASYLR	DSAV	MDSDSVI	PSHQVST	ILAKGAERN 943											
Qy	630	FGFLATRL	FRVFKTED	TQ	GKKWK	CC	LPFSFVIFL	---	---	IIGCIIS	GITLLAIFRV 683											
Db	944	SYRLS	---	WGTE	NLDNAL	SSSPH	SFGFLVIF	FWDARG	GA	MRCR	HNGLII	---I 993										
Qy	684	DPKHLTV	NAVLISIAS	VVGLAF	VNCR	TWQV	LDLSLLNS	ORKL	---	---	---	HNAASKL 734										
Db	994	PPRKCT	AAPT	---	---	RVTC	RL	---	---	VKRH	LATMP	MPVEGEGLASRL 1030										
Qy	735	HKLKSEG	---	---	FMKVL	KCEV	ELMARM	AKTID	SFTQN	---	---	QTLRLV	LIDG 776									
Db	1031	IEVGP	GAQFL	GPVVEI	PHFAAL	R	KGKRLV	VLRS	ENGDS	WKEH	FCDY	TEDE	LEINLG 1090									
Qy	777	LDACEQ	DKVLQ	MLD	TVRVL	FSK	GFPIA	FASDPH	--II	IKAINQ	NLNS	VLDR	DSNINGHDY 834									
Db	1091	MD	---	---	EV	LDSP	EDLEK	KICRIIT	TRDFQY	FAV	VSRIQ	---	---	DSNL	GIPE	1134						
Qy	835	MRNIVHL	VPFLN	SRGLS	NARK	FLVTS	ATNGD	VP	CSDDT	GTIG	QEDAD	RRV	QNS	IGEMTK	-- 892							
Db	1135	---	---	GGVLS	STVWP	QVAV	PEGALTKRI	---	---	RVGLQ	---	---	---	---	---	AQPMHSE	LVKKI 1176					
Qy	893	LGSKTALN	RREDTV	---	RRRQ	WORI	T	QMS	FDLTK	L	LV	TED	W	PSDI	SPQ	MRRLN	IVSVT 950					
Db	1177	LGNKAT	FSPTV	LE	PRRR	K	FKP	HTWTPI	V	PKPASS	DV	MLNG	FCG	DAP	TLRL	---	---	CSIT 1232				
Qy	951	GRLLRANQ	ISFN	DR	LAS	WINL	---	---	---	---	---	---	---	---	---	---	---	---	1001			
Db	1233	GGITPA	---	QWED	ITG	TP	PLIF	VNE	CVS	FTTN	VSAR	FWLI	---	---	DCRQ	I	Q	SV	TFASQ 1283			
Qy	1002	IYERI	---	---	SKNI	PT	KD	VEPLL	---	---	---	---	---	---	---	---	---	---	---	1044		
Db	1284	VYREI	ICVP	YMAK	FPV	FAKSHD	PIEAR	LCF	CMTD	DKV	DKLEQ	QENF	---	---	---	---	---	---	---	1339		
Qy	1045	VKV	---	---	FLP	CTV	NLD	P	---	---	---	---	---	---	---	---	---	---	---	1070		
Db	1340	VEVLE	KPIV	D	CFGN	L	VLPTK	S	GOH	HF	F	FAK	ENR	LP	LVK	V	DR	T	TOEP	CGRLS	FMK 1399	
Qy	1071	EQIS	IG	L	A	P	---	---	---	---	---	---	---	---	---	---	---	---	---	---	1123	
Db	1400	EPK	STR	GLV	HQA	IC	N	L	N	I	L	P	I	---	---	---	---	---	---	---	1448	
Qy	1124	SYIS	G	M	T	G	P	O	H	P	F	N	R	G	S	G	P	A	P	V	V	1183

Db	1449	TETSVL---KSH-----LVNEVPEVLASPDLLSEVSEMKQDLIKM---TAILT	1489
Qy	1184	ANINGRVLACQNIIDELKEMNMFGDMHLFRSTVLEMRNAESHVVPEPRFLSE-SSSGP	1242
Db	1490	TVSDKA-GSIKVELVKAEEEPGE---PPEIVER-----VKEDLEKVNELRSQT	1537
Qy	1243	APHGEPARASHNELPHTELSSQTPYTLNFSPFEELNTLGLDEGAPRHSLNSQSQTRTP	1302
Db	1538	CTRESSVQSSRSERGLVE---EEMVIVSDEIEEAR-----QKAPLEITEYPCVEVRIDK	1590
Qy	1303	SLSSINSODSIEIETSKLTDKQVAEYRDAYREYIAQMSOLEGPGSTTISGRSSPHSTYIM	1362
Db	1591	EIKGKVEKDSGLVNYLTDLLNT-CVPLPKSQLOTVQDKAGKCEALAVGRSS-----	1642
Qy	1363	QSSSSGGSIHNLQEKQKQSEP-----KPDGCKRFLMKRGDVIDYSSSGVSTN	1412
Db	1643	-----EKEGKQIPPETQSTQKHKPSLGIKKPVRRK-----	1674
Qy	1413	DASPLDPITEDEKSDQSGSKLLPQKKSRSLSLFTQDLKLGSGRLYQKLPDEDESGT	1472
Db	1675	---LKEKQKQKEGLQASAEKELKGSSESLGE-----DFGLAPEPLPTVKATS--	1722
Qy	1473	EESNTPLDKDDKQKABKQVVPKSP-----HSABPITFIKAKBYLSDALLDKQDS	1527
Db	1723	-----PLIEETPIGSIKDKVKALQKQKRVEDQKGRSKLPPIR--VKGEDVPPKTTTHRPH	1774
Qy	1528	SDSGVRSE-----SSPNHSLHNEVADDSOLEKANLIELEDDSHSGKRGIPHSLGLQD	1581
Db	1775	ANSPSLKSERIAPGSPKTERHSTLSSAKTERHPVPS-----PSSKTEKHS	1822
Qy	1582	PIIARMSICSDKSPSCSLIASSPEENWPACQKAYNLNRTPTVTTLNNSAPANRANQ	1641
Db	1823	PV---SPSAKTERHSPASS---SKTEKHSFV-----SPSTKTERHSPVSTKTER	1867
Qy	1642	NFDEMEGIRETSQVILRPSSPNPTTQNIENLK-SMTHKRGQRSSYRLSKDPP	1694
Db	1868	H---PPVSPSGTKDKRPVSPSGRTEKHPPVSPGRTEKRLPVSPSGRTDKHQP	1917
RESULT 9			
T42716			
ankyrin 3, splice form 4 - mouse			
C;Species: Mus musculus (house mouse)			
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004			
C;Accession: T42716			
E;Peters, L.L.; John, K.M.; Lu, F.M.; Eicher, E.M.; Higgins, A.; Yialanias, M.; Turtzo, I.			
J. Cell Biol. 130, 313-330, 1995			
A;Title: Ank3 (epithelial ankyrin), a widely distributed new member of the ankyrin gene			
the repeat domain.			
A;Reference number: 222237; PMID:95340633; PMID:7615634			
A;Accession: T42716			
A;Status: preliminary; translated from GB/EMBL/DBJ			
A;Molecule type: mRNA			
A;Residues: 1-1961 <PET>			
A;Cross-references: UNIPROT:Q61307; EMBL:U40632; NID:g710548; PID:g710552; PIDN:AAB01607			
A;Experimental source: strain C57BL/6J; kidney			
C;Genetics:			
A;Gene: Ank3			
A;Map position: 10			
C;Superfamily: ankyrin; ankyrin repeat homology			
C;Keywords: alternative splicing			
Query Match 6.0%; Score 530; DB 2; Length 1961;			
Best Local Similarity 19.9%; Pred. No. 4.7e-21;			
Matches 432; Conservative 279; Mismatches 673; Indels 784; Gaps 90;			
Qy	14	EEENIPALKALKCKVDNERNECGQTPMTAAEQGNLEIVKELIKKGNANLELDLQNT	73
Db	66	KEGHEVVESELQREANVDAATKGNLTALHAGQAEVVKLVLTGANNVAQSNGFT	125
Qy	74	ALISASKEGHVHIYEELKCGNLEHDMGWTALMMAKYGRDVTVVLELLI-----	124
Db	126	PLYMAAENHLEVVRFLLDNGASQSLATEDGFTFLAVALQGHQDVVSLLENTKGVK	185

Qy	125	-----SHGANPSVTGL-----	135
Db	186	LPALHIAARKDDTKAAALLQNDTNADVESKSGTGPLHIAAHYGNINVATLLNRAAAMD	245
Qy	136	---QYSVPIIWAAGRGHADIVHLLONGAKVNSDKYGTTPPLVWAARKG-----	182
Db	246	FTARNDITPLHVASKRGVANNVKLLDRGAKIDAKTRDGLTPLHCGASGHEQVVMELLD	305
Qy	183	-----HLECVKHLAMGADVDDQEGANSMTALIVAVKGGYTQS	219
Db	306	RSAPILSKTKNGLSPLHMATQGDHLCVQLLQHNVPDDVTNDYLTALHVAACHGCHYKV	365
Qy	220	VKELIKENPNVLTDKDGTALMIASKE-----	247
Db	366	AKVLLDKKASNAKALNGFTPLHIAKKNRIRVMELLKKGASTQAVTESGLTPIHVAAF	425
Qy	248	-GHTETIQQDLADAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRGQNKTA	306
Db	426	MGHVNIIVSQLMHHGASPNNTTVRGETALHMAARSGAEVVRYLVQDGAQVEAKAKDDTP	485
Qy	307	LYNAVEKGNATWDDIIOCHPDTEICTKDGETPLIKATKMRNIEVVELLLDKGAKVSAVD	366
Db	486	LHISARLGKADIVQQLLQCGASPNAAATTSYTPPLHAAAREGHEDVAAPFLDHGASLSITT	545
Qy	367	KKGDTPLHIAIRGRSKRLAELLRL-----NPKDGRLL-----	398
Db	546	KKGFTPLHVAKYKLEVASLLLOKSASPDAGKSGLTPLHVAAHYNQKVALLLDQGA	605
Qy	399	--YRPNKAGETPNIDC-SHQKSILTQI--FGAR-----HLSTPTDQDMLGY	441
Db	606	SPHAAAKNGYTPLHIAAKNQMDIATSILEYGADANAVTROGIASVHLAAQGHVDMVSL	665
Qy	442	DLYSALADILSEPTMOP-----PICVGLYAOQ	469
Db	666	LLSRNANVLSNKSGLTFLHAAQEDRVNAEVLVNOGAHVDAQTKMGYTPLHVGCCH--Y	723
Qy	470	GSKG--SFLK--KLEDEMKT-----FAGQQIEPLFQFSWLIVFLTLCCGG--LG	514
Db	724	GNIKIVNPLLOHSAKNAKNGYTAHQAAQ-----GHTHIIN	763
Qy	515	LLFAFTVHPN-----LGIAVLSFLALLYIFFIVI-----	544
Db	764	VLLQNNASPNELTVNGNTALAIARRLGVISVVDTLKVTEIIMTTTTITEKHKNVPETM	823
Qy	545	-----YFGRREGESWNWAVLSTRLARHIGVLELLKUL--	578
Db	824	NEVLDMSDDEVKASAPEKLSDGEYISDGBEGDKCTWFKIPKVQ-----EVLVKS	875
Qy	579	-----MFVNPPPELPTTKALP-----VRFLETD-----YNRLLSSVGG	612
Db	876	AITGDTKYLGPQDLKELGDDSLPAGYVGFSLGARSASURSFSSDRSYTLNRSYARDS	935
Qy	613	TSLAEMIATLSDACEREFGLFATLRFVFKTEDTQGGKKKKKTCCLPSFVIFLFIIGCII	672
Db	936	MMIEELLVP-----SKEHLTTFTEF-----DSDSLRHSWAADT-----LDNV	978
Qy	673	SGI--TLALIFRVDPKHLTVNAVLIISASVGLAFVL--NCRTWQVLDLSLNSQRKL	727
Db	979	SPVHSGFLVSFMVDARGSMRG-----SRHHGMRIIPPRKCTAPTRITCLV--KRHL	1031
Qy	728	HN-----AASKLHKLKSEG-----FMVKLCEVELMARMAKTIDSTQ	765
Db	1032	ANPPMVEGSLASFLVEMGPAGAQFLGPVIVEIPHFGSGMRGERELIVLSENGETWKE	1091
Qy	766	NQ-----TRLVIIDGLDACEQDKVLQMLDTVRV--LFSKGPFTIAIFASDPIIKAINQ	818
Db	1092	HQPKDKNEDLAELNGND--EELDSPELGTGKRICRIITK--DFQYFA-----VVSRIKQ	1143
Qy	819	NLNSVLBDSNINGHDYMRNIVHPLFVFLNSRGLSNARKFLV--TSATNGDVPCSDTTGQ--	875
Db	1144	ESNQIGPEGGI-----LSSTTVPLVQASFPFEGALTKRIRVGLQAO	1183

QY 876 ---EDADRRVQNSIGEMTKLGSKTALNRDVTY--RRQWQRTITROMSFDTLTKLL---V 927
Db 1184 VPPEETVKKI-----LGNKATFSPVIVVEPRRRKPKHPIIT-MTIPVPPSGEGV 1231
QY 928 TEDWFSDISPQTRRLINIVSVTRGLLRANQISFNWDLASMINLQWMP-----Y 978
Db 1232 SNGYKGDATP-NLRLCSITGTSIP-----AQWEDITGTPLTIFKDCVSF 1276
QY 979 RTS-----WL-----ILYLETSGIPQWMLTKI-YERISKNTPTTKDVEPLEIDGDIR 1027
Db 1277 TTNVSARFNLADCHQVL---ETVGLASQLYRELICVPYMAKFFVFAKTNDP---VESSLR 1330
QY 1028 -----NEVFLLSSTPVLVARDVKVFLPCTVNLDPKLK---EIIADYR 1067
Db 1331 CFCWTDNRVDKTLQEQENFEFVARSK-DIEVLEGKPIYDVCYGNLAPLTGKGQQLVFNFY 1389
QY 1068 AAREQISIGLAVPPLPLHEGPPRPSGY-----SOPPSVCSSTSFNGPPAGGWSQPH 1122
Db 1390 SPKEN-----RLPFSIKIRDTSQEP--CGRISP-----1415
QY 1123 SSYYSGMTGQHPFYNRGSPAPGPVVLLNSLNVDAVCEKLQIEGLDQSMPLQYCTTIK 1182
Db 1416 -----LKEPKTKGLPQTAV-----1430
QY 1183 KANINGRVLAQCNIDELKEMNNFGDWHLFRSTVLEMRNAESHVVPEDRPFLESSSGP 1242
Db 1431 -CNLNTLTPAHHKAEKADRQOS-----FASLALRKRY-----YLTEPMSGP 1471
QY 1243 APHGEPAIR-----ASHNELPHTLSQTPYTLNFSPEELNTLGLDEGAPRHNLNLS- 1293
Db 1472 Q---SPCERTDIRMAIVADHLGLSWELARE---LNFVDEINQIRVEN---PNLSIQ 1521
QY 1294 -----W---QSOTRRTPLSLNSQDSISIEISKLTDKVOABY-----RDARYRYIAQ 1337
Db 1522 SFMLLKQWTRDCKNATTDALTSVLTAKINRIDIIVTLLEGPIFDYGNISGRSPADENNVF 1581
QY 1338 MSQLEGGPG-----STTISGRSSPHSTYYMGQSSSGSISHS--N 1374
Db 1582 HDVVDGHPSPQVELETGMYLWTPPNPFQDDHFSDISSESFPRTSPRLSDGLVPSQGN 1641
QY 1375 LEQEGK-----KDS---EPKPDGKRKFLMKRGDVIDYSSSGVSTNDASPLDPTTEED 1424
Db 1642 IEHPTGGPPVTAEDTSLSDKMD-----SVTVDPADPLD--VDES 1682
QY 1425 EKSQSGSKL-----LPKKSSERS--LPQDTLKLGSLYQKLPDSEDESGETE--- 1473
Db 1683 QLKDLQSECAQCWASVPGIPNDGRQAEPLRPQTRKVGMSSEQQEKGKSGPDEVEDTKV 1742
QY 1474 -----ESDNTPLLKDKDKRKAEGKVERVP-----KSPHSAB-PIRTFIK 1512
Db 1743 KSLFEDIQEEVAEEMTEDQGMALNRVORABELAWSLAWQNETPSSGLESFPAQA--- 1799
QY 1513 AKEYLSDALDKDSDSGVRSS-----ESSPNHSLHNEVADDSQLEKANLIE 1560
Db 1800 --RRLTGLLDRLDDSDQARDISYLTGCEPKIEANGNHTA--EVIPEA---KAKPYF 1952
QY 1561 LEDDSHGKRGKI-----PHSLSGIQDPTIARMSCISEDKSPSECS--LIASSPEENW 1611
Db 1853 PESQNDIGKQSIKENLKPCKTHGCGRTEEPV---SPLTAQKSLSEETSKLIVIEDAPKCV 1908
QY 1612 PACQKAYNLNRTSTVTLNNSAPANRANQNFDEMEGIRTSQVILRPSSSPNPTTIONE 1671
Db 1909 FVGWK--KMTRTTAD-----GKARLNLOEPEG-----STRSEPK--QGE 1943
QY 1672 NLKSMTHK 1679
Db 1944 GYKVKTK 1951

RESULT 10

T42715

ankyrin 3, splice form 3 - mouse

C,Species: Mus musculus (house mouse)

C,Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C,Accession: T42715
R,Peters, L.L.; John, K.M.; Lu, F.M.; Eicher, E.M.; Higgins, A.; Vialamas, M.; Turtzo, L.
J. Cell Biol. 130, 313-330, 1995
A,Title: Ank3 (epithelial ankyrin), a widely distributed new member of the ankyrin gene
the repeat domain.
A,Reference number: Z22237, MUID:95340633, PMID:7615634
A,Accession: T42715
A,Status: preliminary; translated from GB/EMBL/DBJ
A,Molecule type: mRNA
A,Residues: 1-1940 <PEP>
A,Cross-references: UNIPROT:O61307; EMBL:L40632; NID:g710548; PID:g710549; PIDN:AAB01604
A,Experimental source: strain C57BL/6J; kidney
C,Genetics:
A,Gene: Ank3
A,Map position: 10
A,Introns: 834/1
A,Superfamily: ankyrin; ankyrin repeat homology
C,Keywords: alternative splicing

Query Match 6.0%; Score 529.5; DB 2; Length 1940;
Best Local Similarity 20.1%; Pred. No. 4.9e-21;
Matches 431; Conservative 280; Mismatches 673; Indels 763; Gaps 90;

QY 14 ESENIPALKALLECKVDNERNECQOTPLMIAAEOGNLEIVKELIKNGANCNLEDLNDWT 73
Db 66 KEGHEVVSSELLQREANVDAATKKGNTALHIASLAGAEBVVKVLTNGANVNAQSONGFT 125
QY 74 ALISASKEGHVHIVEELLKCGVNLHRDMGVTALMWACVKGRTDVVELLL 124
Db 126 PLYMAQENHLEVRFLDLONGASQSLATEDGFTPLAVALQOQHDQVVSLLLENDTKGVR 185
QY 125 -----SHGANPSVTGL----- 135
Db 186 LPALHIAARKDITKAAALLQNDTNADVESKSGFTPLHIAAHYGNINVTALLNRAAAYD 245
QY 136 ---QYSVYPIIWAAGRGHADIIVHLLQNGAKVNCSDKYGTPPLVMAARG- 182
Db 246 FTARNDITPLHVASKRGANVMKLLDRLGAKIDAKTRDGLTPLHCGARGSEHQVEMLLD 305
QY 183 -----HLECVKHLAMGADVQDEGANSMTALIVAVKGYTQS 219
Db 306 RSAPILSKTKNGLSPLHMATQGDHLNCVQLLQHNVPVDVTDNDYLTALHVAACHGHYKV 365
QY 220 VXEILKRNPNVNLTDKGNNTALMIASKE----- 247
Db 366 AKVLLDKKASPNAKALNGFTPLHIAACKNRIRVMELLKHGASIQAVTESGLTPIHVAAF 425
QY 248 -GHTEIVQDLDAGTYVNIPIRDSGDTVLIGAVRGHVEIVRALLOKVADIDIRQDNKTA 306
Db 426 MGHVNIIVSQMHGASPNNTNVRGETALHMAARSGAEEVRYLVQDGAQVEAKAKDQTP 485
QY 307 LYWAVEKGNATVMVRDILQCNPDTEICTKDGETPLIKATKRNIEVVELLLDKGAKVSAVD 366
Db 486 LHISARLGKADIVQQLQOQASPNATTSYTPPLHAAEGHEDVAFLDHGASLSITT 545
QY 367 KGGDTPHIAIRGRSKLAELLR-----NPKDGRLL----- 398
Db 546 KKGFTPLHVAAYKGLVAVASLLIQKSASPDAAKSGSLTPLHVAAYHNDKQVALLDDQGA 605
QY 399 --YRPNKAGETPNVNDK-SHOKSILTQI--FGAR-----HLSPTETDGMGLY 441
Db 606 SPHAAAKNGYTPPLHIAAKNQMDIATSLLEYGADANAVTQGIASVHLAAQEGHVDMSL 665
QY 442 DLYSSALADILSEPTMQP-----PICVGLYAOV 469
Db 666 LLSRNANVNLNKSGLTPLHAAQEDRVNVAEVLNVNQAHAQATKRWGYTFLHVGHCH--Y 723
QY 470 GSKG--SFLKK---KLEDEMT-----PAGQOIEPLFOFSWLIIVFLTLCCG-----LG 514
Db 724 GNIKIVNFFLQHSKAVNAKTNGYTHALHQAQO-----GHTHIN 763
QY 515 LIPAFVTHFN-----LGIAVLSLFLALYIPFIV----- 543

Db 764 VLLQNNASPNELTVNGNTALAIARRGLYISVVDTLKVVTBETIMTTTTITEKHQNVPEM 823
Qy 544 --IYFGRRREGESMNAWVLSTRLARHIGYLELLKL-----MFVNPELPEQTTK 592
Db 824 NEVLDSMDDEGDKCTWFKPKVQ-----EVLKSEDAITGDTDKYLGQDLKELGDD 875
Qy 593 ALP-----VRLFTD-----YNRLSSVGGETSLAEMIAIATSLDACEFGL 633
Db 876 SLPAEYGVGFSIGARSASLRSFSDRSYTLNRSSYARDSMMIEELLVP-----SKEOHLT 930
Qy 634 ATRLFRVPEKTEDTQKKWKCTCLPSVIFLFIIGLIIGI--TLIAIFRVDPKHLTVN 691
Db 931 FTREF-----DSDSLRHYSWAADT-----LDNVLVSSPVHSGFLVSFMDARGGSMR 978
Qy 692 AVLSIASVWGLAFVL-----NCKTWQVLDLSLNSOKRLHN-----AASKLHLKLS 739
Db 979 G-----SRHGWRIIIPRKCTAPTRICRLV--KRHLKLANPPWVEGEGSLARLVEMP 1031
Qy 740 EG-----FMKVLKCEVELMARMARKTIDSTFQNG-----TLVVVIIDGLDACE 781
Db 1032 AGAQLGFPVIVEIPIHFGSMRGKRELIVLRSNGETWKEHQFDSKNEDLAELLNGMD--E 1089
Qy 782 QDKVLQMLDTRV--LPSKGPIAIPASDPHIIKAINQNLNSVLURDSNNGHDMYMRNV 839
Db 1090 ELDSPEELGTKRICRIITK-DPPQVFA-----VWSRIKQESNQIGPEGGI-----1133
Qy 840 HLPVFLNLSRGISNARKFLV--TSATNGDVPDCSDTTGQ-----EDADRVSQNSLGEWTKL 893
Db 1134 -----LSSTTVPLVQSFEGALTKRIRVGLQAQVPVEETVKKI-----L 1173
Qy 894 GSKTALNRDITY--RRRQORTIYQMSFDLTLL-----VTEDWFSDISPQMRLLNIVS 948
Db 1174 GNKATFPIVTEPRRRKFKPIT--WTIPVPPSGEGVSGYKGDATP-NRLLLCSITG 1230
Qy 949 VTGRLLRANQISFNWDRLASINITEOMP-----YRTS-----WL-----ILYLEE 989
Db 1231 GTSP-----AQWEDITGTPLTFIKDCVSTFTNVSGARFWLADCHQVL--E 1273
Qy 990 TEGIPDQMLAKTI-YERISKNIPTTKOVELLEIDGDIR-----NFEV 1031
Db 1274 TVGLASQLYRELICVYNNAKPVVFAKTNDP---VESSLRCFCMTDDRDVKTLEQOENFEE 1330
Qy 1032 FLSSRTPLVARDVKVFLPCTVNLDPKLR---EIIADVRAAREQISIGGLAYPPLPLEHG 1088
Db 1331 VAREK-DIEVLEKPIYVDCYGNLAPLTKGQQLVFNFSYFEN-----1373
Qy 1089 PPRAPSGY-----SQPPSVCSTSFNGFFAGGVSPQPHSSYYSGMTGPQHPFYNRGSGP 1143
Db 1374 --RLPFSIKIRDTSOEP--CGRLSF-----1394
Qy 1144 AGPFWLLNSLVDAVCEKLOIBGLDQSLPQYCTTIKKANINGRVLQACNIDELKEM 1203
Db 1395 -----LKBPTTKGLPQTAV-----CNLNTITLPAHKKAEDARRQ 1429
Qy 1204 NMFGDWHLPRSTVLEMRNAESHVVPEDPRFLSSSSGAPHPGPARP-----ASHN 1255
Db 1430 S-----FASLARKYS-----VLTPFSMSPO--SPCETDIRMAIVADHL 1468
Qy 1256 ELPHTELSSQPTPTLNFSFEELNTLGLDEGAPRHSNLS-----W-----OSQTRRTPSL 1304
Db 1469 GLSWTELARE---LNFVSDBEINQIRVEN---PNSLISQSFMLLKKWVRDGNATTDAL 1521
Qy 1305 SSLNSQSSSIIISKUTDKQAEY-----RAYREYIAQMSQLGGGPG-----1346
Db 1522 TSVLTKINRINDIVTLLEGPIFDYGNISGTRFADENNVPDHPDGFHPSFQVELETPMGLY 1581
Qy 1347 -----STTIGRSSPHETTYMGQSSSGSSTHS--NLQKRG-----KDS---EP 1385
Db 1582 WTPNPPQDDHFDSDISSIESPFRTPSKLSGLVPSQGNIEHPTGGPPVVTAEITSLEDS 1641
Qy 1386 KPDDGRKSFLMKRGVDIVSYSSGVSNTDASPLDPIITEDEKSDQSGSKL-----LFGKK 1439

Db 1642 KMDD-----SVTTDPADPLD--VDSEQLKDLQCSQAQCWASVPGIP 1682
Qy 1440 SSERSS--LFQTDLKLKSGRLRYOKLPSDEDESGTE-----ESDNTPLLLKDDK 1485
Db 1683 NDGRQAEPLRQTRKVGMSSEQQKSGPDEEVTEDEVKSLFEDIQLEEEVAEBEMTDQ 1742
Qy 1486 DRKAEGKVERVP-----KSPHSAB-PIRTFIKAKEYLSDDLKKKSDSDSGVR 1533
Db 1743 QOAMLNVRQAEALAMSSLAGWQNETPSSGLESAPAQ-----RRLTGGLLRLDDSSDQAR 1797
Qy 1534 SS-----ESSPNHSLHNEVADDSOLEKANLIELEDDSHSGKRG-----PH 1574
Db 1798 DSITSYLTGEBGKTEANGNHTA--EVIPEA---KAKPYFPESQNDIGKQSIKENLKPKTH 1852
Qy 1575 SLSLGQPIIARMSICEDKKSPSECS--LIATSPENWPACQKAYNLNRTPTSTVTLNN 1632
Db 1853 GCGRTEFPV-----SPLTAYQKSLSEETSKLVIEDAPKCPVGMK--KMTRTTAD-----1900
Qy 1633 SAPANRANQNPDEMEGIRETSQVILRPPSSSPNPTTIQENILKSMTHK 1679
Db 1901 ---GKARLNLQEBEG-----STRSEPK--QGEYKVKTKK 1930

RESULT 11
149502
Ankyrin - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C:Accession: I49502
R:White, R.A.; Birkenmeier, C.S.; Peters, L.L.; Barker, J.E.; Lux, S.E.
Mamm. Genome 3, 281-285, 1992
A:Title: Murine erythrocyte ankyrin cDNA: Highly conserved regions of the regulatory domain
A:Reference number: I49502; MUID:92345717; PMID:1386265
A:Accession: I49502
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1862 <RES>
A:Cross-references: UNIPROT:Q02357; GB:M84756; NID:g191939; PID:AAA37236.1; PID:g191940
C:Genetics:
A:Gene: Ank-1
C:Superfamily: ankyrin; ankyrin repeat homology
C:Keywords: alternative splicing
F:40-72/Domain: ankyrin repeat homology <AN01>
F:73-105/Domain: ankyrin repeat homology <AN02>
F:106-138/Domain: ankyrin repeat homology <AN03>
F:139-167/Domain: ankyrin repeat homology <AN04>
F:168-200/Domain: ankyrin repeat homology <AN05>
F:201-233/Domain: ankyrin repeat homology <AN06>
F:234-266/Domain: ankyrin repeat homology <AN07>
F:267-299/Domain: ankyrin repeat homology <AN08>
F:300-332/Domain: ankyrin repeat homology <AN09>
F:333-365/Domain: ankyrin repeat homology <AN10>
F:366-398/Domain: ankyrin repeat homology <AN11>
F:399-431/Domain: ankyrin repeat homology <AN12>
F:432-464/Domain: ankyrin repeat homology <AN13>
F:465-497/Domain: ankyrin repeat homology <AN14>
F:498-530/Domain: ankyrin repeat homology <AN15>
F:531-563/Domain: ankyrin repeat homology <AN16>
F:564-596/Domain: ankyrin repeat homology <AN17>
F:597-629/Domain: ankyrin repeat homology <AN18>
F:630-662/Domain: ankyrin repeat homology <AN19>
F:663-695/Domain: ankyrin repeat homology <AN20>
F:696-728/Domain: ankyrin repeat homology <AN21>
F:729-761/Domain: ankyrin repeat homology <AN22>
F:762-794/Domain: ankyrin repeat homology <AN23>

Query Match 5.9%; Score 525; DB 2; Length 1862;
Best Local Similarity 20.1%; Pred. No. 8.2e-21;
Matches 396; Conservative 272; Mismatches 648; Indels 654; Gaps 77;
Qy 21 LKALLEKCKVDNERCEGQTPMLTAAECGNLEIVKELIKGANCNLEBDLNTWTALISASK 80
Db 90 VRELNVNGANVNAQSKGFTPLYMAAQENHLEVVKFLENGANQNVATEDGFTPLAVALQ 149

QY 81 EGHVHIVEELKCGV-----NLEHRDMGCTALMWA 111
Db 150 QGHNVVAHLINVTGKGVLPALHIAARNDDTRTAVALLQNDPNPDVLSKTGFTPLHTA 209
QY 112 CYKRGTDVVELLSHGANSPT-----GLOV-----SV 139
Db 210 AHYENLVAQALLNRGASVNFTPQNGITPLHIAARRGNVIMVRLLLDRGAQIETRTKDEL 269
QY 140 YPIIWAAGRHADIVHLLQLONGAKVCSKYGTTPPLVMAARKGHLBCVHLLAMGADV- 198
Db 270 TPLHCAARNGHVRISBEILLDHGAPIQAKTKNGLSPIHMAAQGDHLLDCVRLLLQYNAEIDD 329
QY 199 -----QEGA-----NSMTALIVAKGGYQTSVKILKR 226
Db 330 ITLDHLTPLVAAHCHHRVAKVLLDKGAKPNRSLNGFTPLHIAKKNHIRVWELLKKT 389
QY 227 NPVNLTDKDNTALMIASKEGHTETVQDILLDAGTYVNIPIDRSGDTVLIGAVRGHVEIV 286
Db 390 GASIDAVTESGLTPLHVASPMGHLPIVKNLLQRGASPNVSNKVEITPLHMAARAGHTEVA 449
QY 287 RALLQKYADIDIRQDNKTALYWAKEGNATWVRDILQCNPDTEICTKQGETPLIKATKM 346
Db 450 KYLLQNKAKANAKADQDPLHCAARIGHTGMVKLLLENGASPNLATTAGTTPHTAARE 509
QY 347 RNIEVVELLLDKAKVSAVDKGDTPHIAIRGRSKLAELLRLNPKDGLLYRPNKAGE 406
Db 510 GHVDTALALLEKASQACMTKGTFTPLHVAAKYGVRLAELLLEHDA-----HPNAAG 563
QY 407 ---TPYNIDCSHQ---KSIL-----TOIFGAR----- 427
Db 564 NGITPLHVAHNNLIDIVKLLPRGSPHSPANNGYTPHIAKQNIQEVARSLLQYGS 623
QY 428 -----HLSPTETDGMGLDYLYSSA-----LADIL 452
Db 624 ANAESVQGVTPPLHAAQEGHTEMVALLSKQANGNLKNSGLTPLHLVSGEHLVADV 683
QY 453 SE--PTMOPPICVG-----LYAQWGSK--SFLLKKLED-EMKTFAGQIEPLFQSWL- 501
Db 684 IRHGVTVDATRMGYTPLHVASHYGNIKLVFLQHQADVNAKTKLG--YSPILHQAQOG 741
QY 502 -IVFLTLLCGGLLFAFVPHN-----LGIAVSLFALLIYIPFIVFYFGRRE 551
Db 742 HTDIVTLLKNG-----ASNEVSSNGTTPLAIAKRLGYISVTDVLKVT----- 786
QY 552 GESNNWVLSRLAHIGYLELLLLKLMFVNPPELPEQT-TKALPVRFPLFTDYNRLSSVG 610
Db 787 ----DETSVVLVSDKHRMSYPETVDBILDVSEDEGDELVGSKA-----ERRDSRDVG 834
QY 611 GETSLAEMIATLSDACEREGFLATRLFR-----VFKTED-TQKKKWKKTCLDS 660
Db 835 BEKELDFVPKLDQVVE-----SPAIPRIPCVTPETVIRSEDBQASKEYDEDSLIFS 888
QY 661 FVIFPLIIGIISGITLLAIFRVDPKHLTVNAVLISIASVVGFLAVLNCKRTMWOVLDSL 720
Db 889 -----SPATESDNI-SPVAPSVHTGLVSP-----WVDARG 919
QY 721 NSQRKRLHNAASKHLKLGSEGMVKLKEVELMARMAKTIDSTFQNTQNTRLVWIIDGLDAC 780
Db 920 GSRGSRHN-----GLRWIP-----PRTCAAPTRITCRLV----- 950
QY 781 EQDKVTLMLDTRVTLVLSKGPPIAFASDPHIIKAINQNLNSVLRNSNNGHDMRN-IV 839
Db 951 ----KPQKLTTPPL-----AEEGLASRIIALGPTGAQFLSPVIV 987
QY 840 HLPVFL-NSRGLSNARKFLVTSATNGDVPSCDTTGTQEDADRVSNSLCEMTKLGSKTA 898
Db 988 EIPHPASHGEG---DRELVLVRENSGV-----WKEHKSRYGSEYLDQILN-GWDSRE 1035
QY 899 LNRDRYRRRQMQRTITRQMSFDLTKLVV---EDWFSIDISPTQMRRLNLSVTVGRLR 955
Db 1036 LGSLELEKRVKRCRIIT--TDFPLFYFVINSRLQCD-YDTIGTGPB-----GSLRSKLVP 1085

QY 956 ANQISFNWRDLASWINLTQWPFYRTSWLILYLEETEGIPQDMTLKTIYERISKNIPTTKD 1015
Db 1086 LVQATPEPNAVTKVKALQ-----AQPVPDELVTKL----- 1118
QY 1016 VEPLLEIDGDIRNFVEFLSRTPVLVARDVKVFLPCTVNLDPKRLRIIADVRAAREQISI 1075
Db 1119 -----GNQATF-----SPI-----VTVEPRRRKFRPI----- 1141
QY 1076 GGLAYPPLPLHEGPPRAPSGYSQPPS---VCSSTSGNFPAGVSPQPHSSYSYSGMTGP 1132
Db 1142 -GLRIPLPSPWTDNPR-DSGEGDTTSLRLCS-----VIGTDOAQ-----WEDITGT 1187
QY 1133 QHPFYNR-----GSGPAPGPVLLNSL---NVDACEKLL-----QIEGL 1169
Db 1188 TKLIYANECANTTNVSARFWLSDCPTABAVHFATLLYKELTAVPYMAKFFVIAKNDA 1247
QY 1170 DOSMLPYQCTTIKK-----ANINGRVLAQCNIDELK 1200
Db 1248 REGRLACYCMTDDKVDKTLLEQHNFVEVARSRDIEVLEGMPLFAELSG-----NLVPVK 1301
QY 1201 KEMNNFQGMHLFRSTVLEMRNAESHVVPEDPRFLSESSGAPAGHGEPAARRASHNE----- 1256
Db 1302 KAAQORSPHFQSFRENRLAI-----PVKVRDSSREPFGFLSLRKTMYEDTOH 1350
QY 1257 -LPHTELS-----SOTPYTLNFSFEELNTLGLDEGAPR-----HS 1290
Db 1351 ILCHLNTMPPCTKGGABDRRTTLPLRLYSILSESRLGFTSDTDRVEMNAVIREHL 1410
QY 1291 NLSWOSQTR-----RTPSLSSLSNQSSSI-----EISKLTDKVOAEYRD 1329
Db 1411 GLSWAELARELOFQSVEDINRIVENPNSLLDQSTALLTLVWDREGENAKM-ENLYTALRN 1469
QY 1330 AYREYIAQMSQLEGEGPGSTTISGRSPHSTTYMGQSSSGSIHNSLEQEKGDSEPKDD 1389
Db 1470 IDRSEIVNM--LEG-----SGRQS-----RNLKPERRH 1495
QY 1390 GRKSLMKRGDVIDYSS-----SGVSTNDASPLDPIITEED 1424
Db 1496 GOREYLSLSPQVNGYSSLODELLSPASLYALPSPLCADOYWNEVTVIDAIPL-AAEYHD 1554
QY 1425 ---EKSDQS--GSKLLPGKKSERSLFT-----DLKKGSLRYQKLP-----SD 1466
Db 1555 TWLEMSDMQVMSAGLTPSLVTAEDSSLECSKAEDSDAIPWKLEGAHSEDTOGPELGSOD 1614
QY 1467 EDESGTEESDNT-----PLLKDDKDKAKGVKERVKPSPHSABPIRTFFAKAYELSDALL 1522
Db 1615 LVDEDDTVDSDATNGLADLLQEGORSEKKQREV-SGTQDTEVSLVSGQQRVHARIT 1673
QY 1523 DKK-----DSSDSGVRSSESSPNHSHNEVADDSQLEKANLIELEDSDSHSGKGI--- 1572
Db 1674 DSPSVQVLDJRQARTLWDKOGSTAVHPOEATQSSWQB-----EVTQGHSHFORRITTIQ 1729
QY 1573 ---PHSLSGLOPFIITARMSICSEDKSPSECSLIASSPEENWPACQAYN 1619
Db 1730 GPPEGALQVEQVIV---STREHVQRGPPEPTGSPKAGKEPSLWAPESAFS 1776

RESULT 12

T42713

C:Species: Mus musculus (house mouse)

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004

C:Accession: T42713

R:Peterson, L.L.; John, K.M.; Lu, F.M.; Eicher, E.M.; Higgins, A.; Vialamas, M.; Turtzo, L.

J. Cell Biol. 130, 313-330, 1995

A:Title: Ank3 (epithelial ankyrin), a widely distributed new member of the ankyrin gene

the repeat domain

A:Reference number: Z22237; MUID:95340633; PMID:7615634

A:Accession: T42713

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1943 <P>

A:Cross-references: UNIPROT:Q61307; EMBL:L40632; NID:g710548; PID:g710550; PIDN:AAB01606

A;Experimental source: strain C57BL/6J; kidney

C;Genetics:

A;Gene: Ank3

A;Map position: 10

A;Introns: 855/1

C;Function:

A;Description: supposed to play an important role in the polarized distribution of many

A;Note: major kidney ankyrin

C;Superfamily: ankyrin; ankyrin repeat homology

C;Keywords: alternative splicing

Query Match 5.9%; Score 523; DB 2; Length 1943;

Best Local Similarity 20.2%; Pred. No. 1.1e-20;

Matches 437; Conservative 270; Mismatches 660; Indels 800; Gaps 90;

```
QY 14 EEENIPALKALECKVDVERNECQOTPLMTAAEGNLEIVKELIKNCANGNLEDLDNWT 73
DB 66 KEGHEVVESELLOREANVDAATKGNLTALHSLASGAQAEVVKVLVTNGANVNAOSQNGT 125
QY 74 ALISASKEGHVHIVBEELKCGVNLHEDRMGWTALMWCYKGRDIDVVELL 124
DB 126 PLYMAAQENHLEVVRFLDNGASQSLATEDGFTPLAVALQQGHDQVVSLLLENDTKGKVR 185
QY 125 -----SHGANPSVTGL----- 135
DB 186 LPALHIAARKDDTKAAALLQNDTNADVESKSGFTPLHIAAHYGNINVAITLLNRAAAYD 245
QY 136 ---QVSVPYIIWAAGRHADIVHLLLONGKVCNSDKYGTTPLVWAARKG----- 182
DB 246 FTARNDIITPLHVASKRGNANVKLLDRGAKIDAKTRDGLTPLHCGARSQHEQVEMLLD 305
QY 183 -----HLECVKHLAMGADVDDQEGANSMTALIVAVKGGYTQS 219
DB 306 RSAPILSKTKNGLSPLHMAATQGDHLNCVQLLQHNVPVDDVNDYLTALHVAACHGKVK 365
QY 220 VKEILKRNPNVNLTKDQNTALMASKE----- 247
DB 366 AKVLLDKKASPNAKALNGFTPLHIAACKNRIRVWELLKHGASIQAVTESGLTPIHVAAF 425
QY 248 -GHEIIVODLDAGTYNIPDRSGDTVLIGAVRGHVEIVRALLQKADIDIRQDNKTA 306
DB 426 MGHVNIIVSQLMHHGASPNNTVRGETAHLMAARSGAEVVRVYVQDGAQVEAKAKDQTP 485
QY 307 LYMAVEKGNATWRDILQCNPDTEICTKDGTEPLIKATKMRNIEVWELLKDGAKVSADV 366
DB 486 LHSARLGKADIVQOLLOQASPNNAATTSGYTPLHAREGHEDVAEFLLDHGLSLITT 545
QY 367 KKGDTPLHIAIRGRSRKLAELLR-----NPKDGRLL----- 398
DB 546 KKGFTPLHVAAYKGLKLEVASLLQKASPDAAKSGSLTPLHVAHYDNQVALLLLDQGA 605
QY 399 --YRPNKAGETPYNIDC-SHOKSILTQI--FGAR-----HLSPTETDGMILGY 441
DB 606 SPHAAKNGYTPHIAAKNQMDIATSLLEYGADANAVTROGIAVHLLAQEGHVDVMSL 665
QY 442 DLYSALADILSEPTMQB-----PICVGLVAQM 469
DB 666 LLSRANVNLSKSLTPLHAAQEDRVNVAEVLVNOGAHVDAQTKMGYTPHVGCH--Y 723
QY 470 GSGK--SFLK--KLEDEMT-----FAGQQ-----IEPLFQFSMLIVFTLLCGGL 513
DB 724 GNIKIVNELLQHSARVNAKTNGYVTHALHQAQQGHTHIIIVLLQNN----- 769
QY 514 GLLFAFTVHPNLGIAVSLFLALLYIFIVIFGRRREGESWNVAWLSTLASHIGYLE 573
DB 770 ASPNELTVNGTALAI-----ARRLGYS 793
QY 574 LLLKLMFY-----NPPELPQTTKALPVRFLETDYNNRLSSVGGETSIAEMIA 620
DB 794 VVDTLKVVTIEIMTTTTITEKKNVNPETMEVLD---MSDDEVKASAPKELSDGEYI- 849
QY 621 TSLDACERE-----FGF-LATR--LFRVFKTED--T 646
```

```
DB 850 --SDGEGEDAITGDTDKYLGPDQLKELGDSLPAEGYVGFSLGARSASLRSFSSDRSYT 907
QY 647 QGKKKWKTKCCLPSFVFLPIIGICIISGITLAIIRVDP-KHLT-----VNAVLIISI 697
DB 908 LNRSSYAR---DSMHEELVPSKEQHLTTRTFDSLSLHYSWAADTLONVNLVSSPV 963
QY 698 ASVVGLAFVLNCRTWQVLDLNLNQRKRLHNAAS-----KLHLK- 737
DB 964 HSGFLVSMVDARG-----GSMGRSRHGMRIIPPRKCTAPTRITCRLVKRHKLA 1014
QY 738 -----KSEG-----FMKVLKCEVELMARMMAKTIDSTFQON 766
DB 1015 NPPPMVEGEGSLASLAVEMGPAGQFLGPVIVEIPHFGSMRGKRELIIVLRSENGETWKEH 1074
QY 767 Q-----TRLVVIIDGLDACEODKVLQMLDTRV--LFSKGPFIASFADPHIIKAINON 819
DB 1075 QFDSKNEDLAELNGMD--BELDSPBELGTVKRICRITK-DPQVFA-----VVSRIKQE 1126
QY 820 LNSVLRSNINGHDMRNIVHLPVFLNSRGLSNARKFLV--TSATNGDVPCSDTTGIIQ--- 875
DB 1127 SNQIGPEGGI-----LSSTTVPLVQASFPEGALTKRIRVGLQAP 1166
QY 876 --EDARRVSONSGEMTKLGSKTALNRDITY--RRROMQRTITRQMSFDLTLLK---VT 928
DB 1167 VPEETVKKI-----LGKATFSPITVTEPRRRKFKHPIT--MTPIPVPPSGEGVS 1214
QY 929 EDWFSDISPQMRLLNIVSVTGRLLRANQISFNWDRLASWINLTEQWP-----YR 979
DB 1215 NGYKGDATP-NLRLLCSITGTSP-----AQWEDITGTTLTFIKDCVST 1259
QY 980 TS-----WL-----ILYLEETEGIPDQWTLKTI-YERISKNIPTTKOVPELLEIDGDIR- 1027
DB 1260 TNVSARFWLADCHQVL--ETVGLASQLYRELIQVPMKPVVFAKTNDP---VSSLRUC 1313
QY 1028 -----NPEVFLSSRTPVLVARDVKVFLPCTVNLDPKLR---EIIADVRA 1068
DB 1314 FCMTDDRVDKTLQEQNEFEVARSK-DIEVLEKPIYVDCYGNLAPLTGKQQLVFNFSY 1372
QY 1069 AREQISIGGLAYPLPLHEGPPRAPSGY-----SQPPSVCSSTSFNGFPAGGVSPQPHS 1123
DB 1373 FKRN-----RUPFSIKRDTSQEP--CORLSF----- 1397
QY 1124 SYSGMTGPQHPFYNGSGPAPGPVLLNSLVNDAVCEKLAQIEGLDQSLMPOYCTTIKK 1183
DB 1398 -----LKEPTKTGKLQTAV----- 1412
QY 1184 ANINGRVLAQCNIDELKEMMNFGDWHLPFSTVLEMRNAESHVVPEDPFLSSSSSSGPA 1243
DB 1413 CNLNIITLPAHKAERKARRQS-----FASLALRKYS-----YLTPSPMSQP 1454
QY 1244 PHGEPARR-----ASHNELPHTLSSQTPYTLNFSFEELNTLGLDEGAPRHSNLS-- 1293
DB 1455 --SPCBERTDIRMAIVADHLGLSWTELARE--LNFSVDEINQIRVEN---PNSLISQS 1504
QY 1294 -----W--QSTRTPTSLSSLNQDSSIBISKLTKVQAEY-----RDAYREYIAQM 1338
DB 1505 FMLLKKVTRDGNKNAATDALTSVLTKINRIDIIVLLESGPIFDYGNISGTRSFADENNVEH 1564
QY 1339 SQLEGGPG-----STTISGRSPSHSTYYMGQSSSGSGSIHS--NL 1375
DB 1565 DPVDGHFSFQVELETPMGLYWTNPPFPQDDHFDSDISIESPFRTPSRSLDGLVPSQNI 1624
QY 1376 EQEKG-----KDS-----EPKPDGGRKFLMKRGDVIDYSSSGVSTNDASPLDPTTEDE 1425
DB 1625 EHPTGGPPVTAEDTSLEDSKDD-----SVTTVDPADLDD--VDESQ 1665
QY 1426 KSDQSGSKL-----LPGKKSSERS--LFQTDLKLXGSLRYQKLPSEDESGTE---- 1473
DB 1666 LKDLQCSECAQCWASVFGIPNDGQAEPLRPQTRKVGMSSEQEKSGKSGPDEEVEDKVK 1725
QY 1474 -----ESDNTPLLKDDKDKAKGKVERVP-----KSPHSAE-PIRTFIKA 1513
DB 1726 SLFEDIQLEVEABEMTEDQQAMLNRVQRAELAMSSLAGWQNETPTSGSLESQAQ---- 1781
```

QY 1514 KEVLSALLDKDSSDGVRS-----ESSPNHSHNEVADDSQLEKANLIEL 1561
Db 1782 -RLTGGLLDRDDSSDQARDSITSYLTGPGRKEANGNHTA--EVIPEA---KAKPYFP 1835
QY 1562 EDDSHSGKRG-----PHSLGLODPIIARMSICSEDKKSPSECS--LIASSPENWP 1612
Db 1836 ESQNDIGKQSIKENLKPHTGCCRTPEV-----SPLTAYQKSLEETSCLVIEDAPKCPV 1891
QY 1613 ACQAYNLANRTPSTVTLNNSAPANRANQNFDEMEGIRETSQVILRPSSPNPTTIQEN 1672
Db 1892 VGMK--KWTRTTAD-----GRARLNQBEEG-----STRSEPK--QGE 1926
QY 1673 LKSMTHK 1679
Db 1927 YKVKTKK 1933

RESULT 13
A55575
ankyrin 3, long splice form - human
N;Alternate names: ankyrin G
C;Species: Homo sapiens (man)
C;Date: 03-Mar-1995 #sequence_revision 03-Mar-1995 #text_change 09-Jul-2004
C;Accession: A55575
R;Kordeli, E.; Lambert, S.; Bennett, V.
J. Biol. Chem. 270, 2352-2359, 1995
A;Title: Ankyrin-G. A new ankyrin gene with neural-specific isoforms localized at the ax
A;Reference number: A55575; MUID:95138209; PMID:7836469
A;Accession: A55575
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-4377 <KOR>
A;Cross-references: UNIPROT:Q12955; GB:U13616; NID:g608024; PIDN:AAA64834.1; PID:g608025
C;Genetics:
A;Gene: GDB:ANK3
A;Cross-references: GDB:424503; OMIM:600465
A;Map position: 10q21-10q21
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
C;Keywords: alternative splicing; peripheral membrane protein
F;73-105/Domain: ankyrin repeat homology <AN01>
F;106-138/Domain: ankyrin repeat homology <AN02>
F;139-171/Domain: ankyrin repeat homology <AN03>
F;172-200/Domain: ankyrin repeat homology <AN04>
F;201-233/Domain: ankyrin repeat homology <AN05>
F;234-266/Domain: ankyrin repeat homology <AN06>
F;267-299/Domain: ankyrin repeat homology <AN07>
F;300-332/Domain: ankyrin repeat homology <AN08>
F;333-365/Domain: ankyrin repeat homology <AN09>
F;366-398/Domain: ankyrin repeat homology <AN10>
F;399-431/Domain: ankyrin repeat homology <AN11>
F;432-464/Domain: ankyrin repeat homology <AN12>
F;465-497/Domain: ankyrin repeat homology <AN13>
F;498-530/Domain: ankyrin repeat homology <AN14>
F;531-563/Domain: ankyrin repeat homology <AN15>
F;564-596/Domain: ankyrin repeat homology <AN16>
F;597-629/Domain: ankyrin repeat homology <AN17>
F;630-662/Domain: ankyrin repeat homology <AN18>
F;663-695/Domain: ankyrin repeat homology <AN19>
F;696-728/Domain: ankyrin repeat homology <AN20>
F;729-761/Domain: ankyrin repeat homology <AN21>
F;762-794/Domain: ankyrin repeat homology <AN22>
F;795-827/Domain: ankyrin repeat homology <AN23>

Query Match 5.9%; Score 521; DB 2; Length 4377; ~
Best Local Similarity 20.6%; Pred. No. 5.6e-20;
Matches 427; Conservative 284; Mismatches 695; Indels 668; Gaps 87;
QY 14 EENIPALKKCKDKVDERNCGQTPMLIAEQGNLEIVKELIRKNGANCNLEDLDNWT 73
Db 83 KEGHVEVSELLQREANVDAAATKKGNTHALIASLAGQAEVVKLVLTNGNVAQAQSQNGFT 142
QY 74 ALISASKEGHVHVEELLKGVNLEHRDGMGTALMWACYKGRDVTVELLL----- 124

Db 143 PLYMAAQENHLEVVVKELLONGASQSLATEDGFTPLAVALQOQHDOVVSLLENDTKKVR 202
QY 125 -----SHGANPSVTGL----- 135
Db 203 LPALHIAARKDDTKAAALLQNDNNADVESKSGFTPLHIAAHYGNINVTALLNRAAAVD 262
QY 136 ---QYVYPIIWAAGRGHADIHILLONGAKVNCSDKYGTTPPLVWAARG----- 182
Db 263 FTARNDDITPLHVASKRGANVMKLLDRGAKIDAKTRDGLTPLHCGARGSEHQVEMLLD 322
QY 183 -----HLECVKHLAMGADVQDEGANSMTALIVAKGVGTOS 219
Db 323 RAAPILSKTYNGLSPLHMATOGDHLNCVQLLQHNVPVDDVTNDYLTALHVAACHYKV 382
QY 220 VKEILKRNPNVNLTKDGNLTALMASKE----- 247
Db 383 AKVLLDKKANPAKALNGFTPLHIAACKNRIKVMELLKHGASIOAVTESGLTPIHVAAF 442
QY 248 -GHEIVQDILLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRQDNKTA 306
Db 443 MGHVNIVSQLMHGASPNNTNVRGETALHMAARSGQAEVVRVLVODGAQVEAKAKDDTP 502
QY 307 LYWAVEKGNATVVRDILQ--CNPD-----EICT 333
Db 503 LHSARLGKADIVQQLQOCASPNATTSYTPHLHSAREGHEDVAFLDHGASISITT 562
QY 334 KDGETPLIKATMRNIEVVVELLDKGAQSVADKGGDTPPLHIAIRGRSRKLAELLRNPK 393
Db 563 KKGFTPLHVAAYKGLVANLLOKASPDAAAGSKSLTPLHVAAHYDNQKVALLLL--D 619
QY 394 DGBLLYRPNKAGETPNIDC--SHQSKILTQI--FGAR-----HLSFTEDGDM 438
Db 620 QGASPHAAAKNGYTPHLHIAAKNQMDIATLLEYGADANAVTRQGIASVHLAAQEGHVD 679
QY 439 LGYDLVSSALADILSEPTWOP-----PICVGLY 466
Db 680 VSLILGRNANVLNKSNGSLTPLHAAQEDRVNVAEVLVNOGAHVDAQTQNGYTPHVGCH 739
QY 467 AQWGSCK--SFLIK-KLEDEMTFAQOQIEPLQFSWLIPLVTLLLCGG-----LGILLF 519
Db 740 --YGNIKIVNLLQHSKAVNAKTNG--YTPLHQAQ-----QGTHIINVLQN 785
QY 520 TWHPN-----LGIASLSFLALLYIFFIYFGRREGESWNAWVLSRLARH-I 569
Db 786 NASPNELTVNGNTALGIARRLGYISVVDTLKIVT-----EETMTTITVTEKHKM 834
QY 570 GYLELLKLKMFVNPPEL-----PEQTTKALPVRFLFTDYNRLSSV--GGTSLA----- 616
Db 835 NVPETWNEVLDDSDDEVKANAPE-----MLSDGEYISDVEGEDAMTGDTKYL 884
QY 617 --EMIAATLSDACEREFGLATRL-----FRVKTED--TOGKKKWKTCCLPSFVIFLF 666
Db 885 GPQDLKELGDDSLPABGYMGFSLGARSASLSRFSDDRSYTLNRSSVAV--DSMMIEEL 940
QY 667 IIGCIISGITLAIKFRVDP-KHLT-----VNAVLSIASVUGLAFVLNCTRWQVLD 717
Db 941 LVPSKEQHLTFTREEDSDSLRHSWAADTLDNVLNVLSSPIHSGFLVSFMVDARG----- 994
QY 718 SLINSORKEHLHNAASKHLKXSEGFMKVL---KCEVELMARMAKTIIDSTQNTQRLAVI 773
Db 995 ---GSVRGSRHNG-----NRIIPRKCTAP-----TRITCLRV-- 1025
QY 774 IDGLDACEQDKVLQMLDVTVRVLFSGPFIATFASDPHIIKAINQNLSVLRDSNINGHD 833
Db 1026 -----KXHLAN-----PPDH-----GERRGISRLVEMGPAGA 1055
QY 834 YMRN-IVHLPVFLNRLSRLNARKFLVTSATNGDVPCSDTTGTQEDADRRVSQNSLEMTK 892
Db 1056 FLGPVIVEIPHFSGMRG--KERELIVLRSENGE-----TWKEHQFD---SKNE--DLTE 1102
QY 893 L--GSKTALNRDVTYRRRQMQRTITROMSFDLTLLVTEDFWSDISFPQWRLNLNVSVT 950

Db 1103 LINGMBELDSEELGKRIICRIITKOF-----POYFAVVS--RIKQSNQIGPE 1150
Qy 951 GRLLRAN-----QISFNWDRILASWINLQTEWPYRTSWLILYEETEGIPDQWTLKTIYER 1005
Db 1151 GGLISSTVPLVQASFPFGALTKIRVGLQ-----AQVPDEIVVKILGNK 1196
Qy 1006 ISKNIPPTKQVEPLEIDGDIRNEVFUSRTPLVARDVKVFLPCTVNLNPKLREIAD 1065
Db 1197 AT-----FSPITVPEPRRKHKPIITMIPV-----1222
Qy 1066 VRAAREQISIGGLAYPLPLHEGPPRAPSGYQSP--PSVCSTSPNGPFAGGVSPQPHS 1123
Db 1223 -----PPSSEGVNSGYKGDTPNLRLLCSIT-----GGTSP-----1254
Qy 1124 SYVSQMTGPHQPFYMRGSGPAPGVVLN-----SLNVD-----AVCEKLKQIEGLQDSM 1173
Db 1255 AQWEDIIGTT-----PLTFIKDCVSTTNVSARFPAWADCHQVLETGLATQL 1301
Qy 1174 -----LPOYCTTIKANINGRVLAQCN-----DELKK--EMNNFQDWHLFRSTVLEM 1220
Db 1302 YRELICVPMKPVVFAKNDPVESLRCFCWTDKVDKTLQEQENFEE-----VA 1352
Qy 1221 RNAESHVPEPRFLESSESSGAPAGEPARA-----SHNELPHT-----ELSSQTPYT 1269
Db 1353 RSKDIEVLEGPIYV--DCYGNLAPLTGKGQQLVFNFYFKEKNRPLFPKIKIRDTSQPCGR 1411
Qy 1270 LNFSGFEELNTLGLDEGAPRHSNLWSQTRTPSLSSLNQSDSIETSKLTKVQAEYRD 1329
Db 1412 LSLFKERTYTKGLQTAVCNINILPAHKKETES-----DQDDEIE--KTRDQSPASL 1463
Qy 1330 AYRE---YIAQMSOLEGGPGSTTIGRSSPHSTY-----YMGQSSGGSIHNSLE 1376
Db 1464 ALRKRYSLTEPFGMIERTGTAT---RSLP--TTYSKPPFPSTRPYQSWTTAPITV-----1513
Qy 1377 QEKGDSEPKDDGRKSLMRKGVVIDYSSSGVSTNDASPLDPI-----TEDEKSDQGS 1432
Db 1514 -----PQPAKSGF-----TSLSSSSSNTPSASPLKSIWSVSTPSPIKSTLGAS 1556
Qy 1433 KLLPKKSSERSLFOFDKLKLGGLR---YQKLPSDEDESGT-----EESDNTPL-----1480
Db 1557 TTSSVKSISDVASPIRS--LRTWSSPIKTVVSQSPYNTQVSGTGLARAPVATEPLKGLA 1615
Qy 1481 -----LKDKDRKAEGKVER-----VPKSPPE-----HSAEPITRFI--KAKELYSDAL 1521
Db 1616 SNSTFSSRTSPVTTAGSLRLSSITMTWTPASPKNINMYSSSLPFPKSIITSAAPLISSPL 1675
Qy 1522 -----LDKDDSDSGVRSESSPNHSL--HNEVA-----DDSOLEKAN--LIE-- 1560
Db 1676 KSVSPVKSRVDVIVSSAKITWASSLSPPVKQMPGHAELVNGSISPLKYASSSTLINGC 1735
Qy 1561 -----LEDDSHSGKGIPIHSLGLQDPTIARMSICSEDKKSPSCSLIASSPEENWPAQC 1615
Db 1736 KATATLQEKISSATNSVSVSAATDVKEKVFSTTTAMPFPLR--SVVSAP-----1786
Qy 1616 KAYNLNRTPTVTTLNNSAPANRANQNFDEMEGIRETSQVILRPSS-----PNPTTION 1670
Db 1787 SAFQSLRTPSASALVTSIGSSISATTS-----SVTSSIIITVPVSVVNVLPALPKKL 1839
Qy 1671 ENLKMTHKRSQSS--YTRLSKDDPPELHAASS 1702
Db 1840 PDSNSFTKSAALLSPIKTLTTTTHPOPHFSRTS 1873

RESULT 14
T42714
ankyrin 3, splice form 2 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C:Accession: T42714
R:Peterson, L.L.; John, K.M.; Lu, F.M.; Eicher, E.M.; Higgins, A.; Yialamas, M.; Turtzo, I.
J. Cell Biol. 130, 313-330, 1995
A:Title: Ank3 (epithelial ankyrin), a widely distributed new member of the ankyrin gene
the repeat domain.

A:Reference number: Z22237; MUID:95340633; PMID:7615634
A:Accession: T42714

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1765 <PET>

A:Cross-references: UNIPROT:Q61307; EMBL:L40632; NID:g710548; PID:g710551; PIDN:AA01605

A:Experimental source: strain C57BL/6J; kidney

C:Genetics:

A:Gene: Ank3

A:Map position: 10

A:Introns: 1587/1

C:Superfamily: ankyrin; ankyrin repeat homology

C:Keywords: alternative splicing

Query Match 5.8%; Score 516; DB 2; Length 1765;

Best Local Similarity 19.8%; Pred. No. 2.4e-20;

Matches 408; Conservative 253; Mismatches 636; Indels 762; Gaps 83;

Qy 14 EEENIPALKLEKCKDVDERNECQOTPLMTAAEQGNLEIVKELIKGANGCNLEDLQNW 73

Db 66 KEGHVEVVSSELLQREANVDAATKKGNTALHTASLAGQAEVVKVLVTNGANVNAQSQNGFT 125

Qy 74 ALISASKEGVHIVVEELLKCGVNLHEDMGWNTALMACYKGRDVTVELLL-----124

Db 126 PLYMAAQENHLEVVRFLDNGASQSLATEDGFTPLAVALQQHQVWSLLLENDTKGVR 185

Qy 125 -----SHGANPSVTGI-----135

Db 186 LPALHIAARKDDTKAAALLLQNDTNADVSKSGGTPPLHIAAHYGNINVTALLNRAA 245

Qy 136 ---QYSYVPIIWAAGRGHADIHLLQNGAKVNSDKYGTTPPLVMAARKG-----182

Db 246 FTARNDDITPLHVASKRGKNANVKKLLDORGAKIDAKTRDGLTFLHCGARSQHEQVVEMLD 305

Qy 183 -----HLCVKKHLLAGADVQEGANSMTALIVAVKGYTQS 219

Db 306 RSAPILSKTKNGLSPLHMATQGDHLCVQLLQHNVPDVTNDYLTALHVAHCHGYKV 365

Qy 220 VKELKENPNVNLTKDGNLTALMTASKE-----247

Db 366 AKVLLDKKASPNKALNGFTPLHIAACKNIRVWELLKHGASIQAVTESGLTIHVAAF 425

Qy 248 -GHTEIVQDLIDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRQDNKTA 306

Db 426 MGHVNIYSQLMHGGASPNNTNVRGETALHMAARSGQAEVVRVYLVDQGAQVEAKAKDQTP 485

Qy 307 LYNAVEKGNATWRDILOCHNPDEICTKDGETPLIKATKMRNIEVVELLLDKGAKVSAVD 366

Db 486 LHISARLGKADIVQQLLQOGASPNAAATSGYTPPLHIAAREGHEDVAAFLLDHGASLSITT 545

Qy 367 KKGDTPLHIAIRGSRKLAELLRNPKDGRLLYRPNKAGETPPYNI DCSHOKSILTOFGA 426

Db 546 KKGFTPLHVAARKVLEVASILLQKSAS-----PDAAG-----KSGLTPDHVA 588

Qy 427 RHLSPTEITDGMGLYDILYSALADILSEPTWQPPICVGLYAQMGSGKSFLLKLEDEMK 486

Db 589 AH-----YDNQKVAL-----LLLDQASPHAAK-----612

Qy 487 FAGQIBPLFQFQSWLIVFLTLCCGGLLPFAFTVHNHNLGIAVLSLFLALIYFVIYF 546

Db 613 -----NGYTPLHIAAKNQMDIATSL-----LEY 636

Qy 547 GG-----RREGESNNAWVLSTRLARHIGYLELLKLMLFVNPPELPBQTTKALPVRFLT 601

Db 637 GADANAVTRQG-----IASVHLAAQEGHVDWVMSLLSRN-----A 671

Qy 602 DYNRLSSVGGETSIAEMIATLSDACEREFGLTFLRFRFKTE-----DTQGGKKW 652

Db 672 NVN-LSNKSGLTLP-----HLAAQEDRVNVAEVLVNOGAHVDAQTQMGY 714

Qy 653 KKTCCLPBSFVIFITGICISGTTLLAIFRVDPKHLTVNAVLIASVGLAFVLCRTW 712

Db 715 TP-----LHVCHYGNIKI-----VNFLLQHSAKV-----NAKT- 743

QY 199 -----NSMTALIVAVKGYTQSVKEILKR 226
Db 338 ITLOHITPLHVAACHGHRVAKVLLDKCAKPNRSLNGFTPLHIAACKKNHVRWELLKKT 397
QY 227 NPNVNLTKDGNATMIASKEGHEIIVODLLDAGTYVNIPIRSGDTVLIGAVRGHVEIV 286
Db 398 GASIDAVTESGLTPLHVASFNGHLPFIVKNLLQRGASPNVSVKVEITPLHMAARAGHTEVA 457
QY 287 RALLQKYADIIRGQDNKTALYVAVEKGNATMVRDILQCNPDTEICTKDGTEPLIKATKM 346
Db 458 KYLLQNKAKANAKKDDOTPHCAARIIGHTGVKLLLENGASPNLATTAGHTPLHTAARE 517
QY 347 RNIEVELLLDKGAKVSADVKGTDPHIAIRGRSRKLAELLNRPDKGRLLYRPNKAGE 406
Db 518 GHVDTALALLEKEASQACMTKGTPLHVAAYKGVRLAELLEHA-----HPNAAGK 571
QY 407 ---TPYNDCHQ-----KSIL-----TQIFGAR----- 427
Db 572 NGLTPLHVAVHHNLLDIVKLLPRGSGSPHSPAWNNGYTPHIAAKQNCQIEVARSLLOYGGS 631
QY 428 -----HLSPTETDGMGLGYDLYSSALADILSEPTMQP-----PIC--- 462
Db 632 ANAESVQGVTPHLAAQEGHTEMVALLSKQANGNKGSLTPLHLVSOEGHVPVADVL 691
QY 463 -----VGLYAQMGSGK--SFLAKKLED-EMKTPAGQOIPLPQFWSL- 501
Db 692 IKHGVTVDAATRMGYTPHVAASHYGNIKLVKFLHQHQAQVNAKTKLG--YSPLHQAQOQ 749
QY 502 -IVFLLTLLCGGLGLLFAFTVHPN-----LGIAVSLFLALLYFFIVFYVGRRE 551
Db 750 HTDIVTLLKNG-----ASPNEVSSNGTTPLAIAKRLGYISVTDVLKVV- 794
QY 552 GESNNAWVLSRLARHIGYLELLKLMFVNPPPEQTITKALPV--RPLFTDYNRLSS 608
Db 795 ----DETSVLVSKHMSYPETVDEILDVS-----EDEGTAHISIMGDELVGSKAERDS 846
QY 609 --VGGETSLAEMIAATLSDACERERFGFLATRFR-----VFKTED-TGKKKKWKTG 656
Db 847 RDVGEERELLDFVKLDQVVE-----SPAIPRIPCVTPTVIRSEDOEASKEYEDS 900
QY 657 CLPSFVIFLFIIGIITLAIIFRVDPKHLTVNAVLIISIASVGLAFVLCNCTRWQVL 716
Db 901 LIPS-----SPATETSDNI-SPVASPVHTGFLVSF-----MV 931
QY 717 DSLNSOKRHLHNAASKLHLKSGFMKVLKCEVELMARMAKTIDSPFQNTQTRVLVUIDG 776
Db 932 DARGSMRSGSRHN-----GLRVIP-----PRTCAAPTRITCRLV----- 966
QY 777 LDACEQDKVLQMLDTRVFLSKGPFPIAFASDPHIIIIKAINONLNSVLRDSNINGHDYMR 836
Db 967 -----KPQKANTPPPL-----AEEGLASRIIAGPTGAQFLS 999
QY 837 N-IVHLPVFL-NSRGLSNARKFLVTSATNGVPCSDTTGIEDADRRVSONSLGEMTKLG 894
Db 1000 PVIIEIPHAFASHGRG---DRELVLVRSNGSV-----WKEHKSRYGESYLDQILN-G 1047
QY 895 SKTALNRDTRRRQMORTIIFROMSFDLTLLVT---EDWFSDISPQTMRELLNIVSVTG 951
Db 1048 MDEBLGSLLEELEKRVCKRIIIT--TDFPLYFVIMRSLCOQD-YDTIGPEG-----GSLRS 1097
QY 952 RLLRANQISFNWDRLASWINLITEQWPYRTSMILYLBETEGIPQOMTLKTIYERISKNIPI 1011
Db 1098 KLVPVQATFENAVTNKVKLALQ-----AQVPDELVTKLL----- 1134
QY 1012 TTQDVEPLEIDGDIRNEFVELSSRTPLVARDVKVFLPCTVNLDPKLRLEIADVRAARE 1071
Db 1135 -----GNOATF-----SPI-----VTVEPRRKRPHRI----- 1157
QY 1072 QISIGGLAYPLPLHEGPAPRAGSYQSPS---VCSSTSFGNGPPAGVSVOPHSSYSYG 1128
Db 1158 -----GLRIPLPSPWTONPR-DSGEGDTTSLRLCS-----VIGGTDQAQ-----WED 1199

QY 1129 MTGPQHFFYNR-----GSGPAPGPVWLLNSL---NVDVAVCEKJK-----Q 1165
Db 1200 ITGTTKLIYANECANFTTUSARFWLSDCPTAEAVHFATLLYKELTAVPYMAKPVIFAK 1259
QY 1166 IEGLDQSMPLPOYCTTIKK-----ANINGRVLAQCN 1196
Db 1260 MNDAREGLRCYCMTDDKDKVKTLEQHENFVEVARSRDIEVLEGMPLFAELSG-----NL 1313
QY 1197 DELAKEMNMFQDWHLFRSTVLEMRNAESHVVVPEDPRFLSESSSGPAPHGEPARRASHNE 1256
Db 1314 VPVKAQAQSRFHFQSFRENRLAI-----PVKVRSSREPFGGFLFLKRTMYE 1362
QY 1257 -----LPHTELS-----SQPYTLNFSFEELNTLGLDEGAPR----- 1288
Db 1363 DTQHILCHLNTMPPCTKSGAEDRRRTLTPLTLRYSLTSSRLGFTSDTORVEMRMABI 1422
QY 1289 --HSNLSWQSOTR-----RTPSLSSINSQDSSI-----EISKLTDKVOA 1325
Db 1423 REHLGLSMAELARELOFVEDINIRVENPNNSLLDQSTALLTLWVDREGENAKM-ENLYT 1481
QY 1326 EYRDAYREYIAQMSQLEGPGSTTISGRSSPHSTYYMQSSSGGSIHNSLNQEKGDSEP 1385
Db 1482 ALNRIDRSEIVNMLE-----VSGRQS-----RNKLP 1507
QY 1386 KPDDGRKSFMLKRGDVIDYSS-----SGVSTNDASPLDPITEDEKSDQSGSKLLPGKK 1439
Db 1508 ERRHGDREYSLSPSQVNGYSSLODELLSPASLOVALP-SPICADQYWNEVTVIDAIPLAA 1566
QY 1440 SSERSSLFQTDLKLKGLRYQKLPSEDE---SGTESDNTPLLLKDDKDKRAEGKVERV 1496
Db 1567 TEHDTMLEMSDMQVMSAGLTPSLVTABDSSLECSKAEDSDAIP-----EWKLEGAHSED 1620
QY 1497 PKSPHSAEPI---RTFIKAYEVLSDALLDKK-----DSSDSGVRSSSES 1537
Db 1621 TQGPDELGSQDLVEDDVTVDSDATNGLADLLGQORVHARITDPSVRQVLDLRSQARTLDWDX 1680
QY 1538 SPNHSLENEVADDSOLEKANLIELEDDSHSGKRG---PHSLSGLODPIIARMISIC 1590
Db 1681 QGSTAVHPQEARQSSWOE---EVTQGHPSFORRITTIQGPPEPGALQEYQVLV---STR 1733
QY 1591 SEDKKSSECSLSLIASSPEENWPACQKAYN 1619
Db 1734 EHVQGGPPETGSPKAGKEPSLWAPESAFS 1762

Search completed: February 5, 2005, 19:28:18
Job time : 58 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 5, 2005, 19:17:40 ; Search time 188.5 Seconds
(without alignments)
4658.970 Million cell updates/sec

Title: US-10-021-571-4

Perfect score: 8884

Sequence: 1 MSVLISQSVINYVEENIPA.....LHAAASSTGFGGERESIL 1715

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8830	99.4	1777	2 Q9ULH0	Q9ulh0 homo sapien
2	8362.5	94.1	1762	2 Q9EOG6	Q9eqg6 rattus norv
3	8266	93.0	1715	2 Q9ERD4	Q9erd4 rattus norv
4	7771	87.5	1693	2 Q80TC7	Q80tc7 mus musculu
5	7567	85.2	1554	2 Q80SX9	Q80sx9 mus musculu
6	6289.5	70.8	1680	2 Q7T163	Q7t163 brachydanio
7	6275.5	70.6	1672	2 Q6P7V1	Q6p7v1 homo sapien
8	5205.5	58.6	1031	2 Q9UF42	Q9uf42 homo sapien
9	3583	40.3	691	2 Q7Z322	Q7z322 homo sapien
10	3211	36.1	664	2 Q6MZU2	Q6mzu2 homo sapien
11	2791	31.4	543	2 Q9H9E4	Q9h9e4 homo sapien
12	2417	27.2	1604	2 Q7KVP5	Q7kvp5 drosophila
13	2417	27.2	1626	2 Q7KVP6	Q7kvp6 drosophila
14	2417	27.2	1678	2 Q9W210	Q9w210 drosophila
15	2328.5	26.2	1354	2 Q7PZV2	Q7pzv2 anopheles g
16	1881	21.2	1089	2 Q7YU92	Q7yu92 drosophila
17	1832.5	20.6	1398	2 Q20109	Q20109 caenorhabdi
18	1255	14.1	246	2 Q9NT37	Q9nt37 homo sapien
19	1119.5	12.6	239	2 Q8K0P2	Q8k0p2 mus musculu
20	704	7.9	257	2 Q7QMP9	Q7qmp9 anopheles g
21	666	7.5	129	2 Q9H889	Q9h889 homo sapien
22	557.5	6.3	1856	2 Q99407	Q99407 homo sapien
23	555.5	6.3	1880	1 ANK1_HUMAN	P16157 homo sapien
24	547.5	6.2	1280	2 Q6N064	Q6n064 homo sapien
25	544	6.1	426	2 Q8YTG9	Q8ytg9 anabaena sp
26	539.5	6.1	1375	2 Q6ZS66	Q6zse6 homo sapien
27	539	6.1	3924	1 ANK2_HUMAN	Q01484 homo sapien
28	525	5.9	1862	1 ANK1_MOUSE	Q02357 mus musculu
29	523	5.9	1863	2 Q7Z315	Q7z315 homo sapien
30	523	5.9	1943	2 Q61307	Q61307 mus musculu
31	522.5	5.9	1726	2 Q8VC68	Q8vc68 mus musculu

RESULT 1

ID	Q9ULH0	PRELIMINARY;	PRT; 1777 AA.
AC	Q9ULH0;		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-OCT-2001 (TrEMBLrel. 18, Last sequence update)		
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)		
DE	KIAA1250 protein (Fragment).		
GN	Name=KIAA1250;		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Brain;		
RX	MEDLINE=20039619; PubMed=10574462;		
RA	Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;		
RT	"Prediction of the coding sequences of unidentified human genes. XV.		
RT	The complete sequences of 100 new cDNA clones from brain which code		
RT	for large proteins in vitro."		
RL	DNA Res. 6:337-345(1999).		
DR	EMBL; AB033076; BAA86564.2; -.		
DR	HSP; P42773; IMX6.		
DR	InterPro; IPR002110; ANK.		
DR	Pfam; PF00023; Ank; 11.		
DR	PRINTS; PR01415; ANKYRIN.		
DR	SMART; SM00248; ANK; 11.		
DR	PROSITE; PS50088; ANK_REPEAT; 10.		
DR	PROSITE; PS50297; ANK_REPEAT_REGION; 1.		
KW	ANK repeat.		
FT	NON TER		
SQ	SEQUENCE 1777 AA; 197209 MW; B6505923FB45F143 CRC64;		

Query Match 99.4%; Score 8830; DB 2; Length 1777;

Best Local Similarity 96.7%; Pred. No. 0;

Matches 1714; Conservative 0; Mismatches 0; Indels 58; Gaps 2;

Qy	1	MSVLISQSVINYVEENIPAKALLEKCDVDNERNECGQTPMTAAEQGNLEIVKELIKN	60
Db	7	MSVLISQSVINYVEENIPAKALLEKCDVDNERNECGQTPMTAAEQGNLEIVKELIKN	66
Qy	61	GANCNLEDLONWTALISASKEGHVHVEELLKCGVNLEHEDMGWGTALMWACYKGRDGV	120
Db	67	GANCNLEDLONWTALISASKEGHVHVEELLKCGVNLEHEDMGWGTALMWACYKGRDGV	126
Qy	121	ELLISHGANPSVTGLQYVPIIWAAGRHADIVHLLQNGAKVNCSDKYGTTPLVWAAR	180
Db	127	ELLISHGANPSVTGL-QYVYPIIWAAGRHADIVHLLQNGAKVNCSDKYGTTPLVWAAR	185
Qy	181	KGHLECVKHLAMGADVDEGANSMTALIVAVKGGYTSQVKEILKRPVNLTDKGNNTA	240
Db	186	KGHLECVKHLAMGADVDEGANSMTALIVAVKGGYTSQVKEILKRPVNLTDKGNNTA	245

QY 241 LMIASKEGHEITEIVQDLIDAGTYVNI PDRSGDTVLIGAVRGCHVEIVRALLQKVDIDIRG 300
DB 246 LMIASKEGHEITEIVQDLIDAGTYVNI PDRSGDTVLIGAVRGCHVEIVRALLQKVDIDIRG 305
QY 301 QDNKTALYWAVERKGNATWVRDILQCNPDTEICTKDGSETPLIKATKMRNIEVVELLDKGA 360
DB 306 QDNKTALYWAVERKGNATWVRDILQCNPDTEICTKDGSETPLIKATKMRNIEVVELLDKGA 365
QY 361 KVSADVKKGTPTLHAIARGSRKLAELLNPNKDGRLLYRPNKAGETPPYNDICSHOKSIL 420
DB 366 KVSADVKKGTPTLHAIARGSRKLAELLNPNKDGRLLYRPNKAGETPPYNDICSHOKSIL 425
QY 421 TQIFGARHLSPTTGDGMLGYDLYSSALADILSEPTWQPPICVGLYAQMGSGKSFLLKLL 480
DB 426 TQIFGARHLSPTTGDGMLGYDLYSSALADILSEPTWQPPICVGLYAQMGSGKSFLLKLL 485
QY 481 EDEMKTIFAGQIQEPLFQFQSWLIVFLTLILCGGLGLLFAFTVHNLGIAVSLFLLALYIF 540
DB 486 EDEMKTIFAGQIQEPLFQFQSWLIVFLTLILCGGLGLLFAFTVHNLGIAVSLFLLALYIF 545
QY 541 FIVYFGRRREGESWNAWVLSRLARHIGYBELLKLMFVNPPELPEQTTKALPVRFIF 600
DB 546 FIVYFGRRREGESWNAWVLSRLARHIGYBELLKLMFVNPPELPEQTTKALPVRFIF 605
QY 601 TDYNRLSSVGGETSLEAMIAATLSDACEREFGLATRLFRVFKTEDTQGGKWKKTCLPS 660
DB 606 TDYNRLSSVGGETSLEAMIAATLSDACEREFGLATRLFRVFKTEDTQGGKWKKTCLPS 665
QY 661 FVIFLFIIGCIISGITLLAIFRVDPRKLTAVNAVLISIASVVGFLAVLNCKTMMQVLDLSLL 720
DB 666 FVIFLFIIGCIISGITLLAIFRVDPRKLTAVNAVLISIASVVGFLAVLNCKTMMQVLDLSLL 725
QY 721 NSQKRLHNAASKHLKKBSEGMKVLKCEVELMARMAKTIDTSFTQNTQRLVVIIDGLDAC 780
DB 726 NSQKRLHNAASKHLKKBSEGMKVLKCEVELMARMAKTIDTSFTQNTQRLVVIIDGLDAC 785
QY 781 EQDKVLQMLDVTVLVSKGPIAIFASDPHIIKATNONLNSVLRDSNNGHDMYRNIVH 840
DB 786 EQDKVLQMLDVTVLVSKGPIAIFASDPHIIKATNONLNSVLRDSNNGHDMYRNIVH 845
QY 841 LPVFLNSRGLSNARKFLVTSATNGDVPSCDITGIEDADRVRVQNSLIGEMTKLGSKTALN 900
DB 846 LPVFLNSRGLSNARKFLVTSATNGDVPSCDITGIEDADRVRVQNSLIGEMTKLGSKTALN 905
QY 901 RRDYTRRRQMRITRQMSFDLTKLVTEWDFSDISFQTMRRLLNIVSVTGRLLRANQIS 960
DB 906 RRDYTRRRQMRITRQMSFDLTKLVTEWDFSDISFQTMRRLLNIVSVTGRLLRANQIS 965
QY 961 FNDRLASWNLTEQWPIRYSMLILVLEETEGIPDQMTLKTIVYERISKNIPTTKDVEPLL 1020
DB 966 FNDRLASWNLTEQWPIRYSMLILVLEETEGIPDQMTLKTIVYERISKNIPTTKDVEPLL 1025
QY 1021 EIDGDIRNFVFLSSRTPVLVARDVKVFLPCTVNLDPKLEIITADVRAAREQISIGGLAY 1080
DB 1026 EIDGDIRNFVFLSSRTPVLVARDVKVFLPCTVNLDPKLEIITADVRAAREQISIGGLAY 1085
QY 1081 PPLPLHEGPPRAPSGVSPVSCSTSFNGPPAGGVVSPQPHSSYSGMTGPOHPYFN 1138
DB 1086 PPLPLHEGPPRAPSGVSPVSCSTSFNGPPAGGVVSPQPHSSYSGMTGPOHPYFN 1145
QY 1139 -----RGSGP 1143
DB 1146 FPAPLYTPRYPGGQHLISRSVKTSLPRQNGLEVIKEDAAEGLSPDTSRGSGP 1205
QY 1144 APGPVLLNSLNDVACEKQLQIEGLDQSMLOPYCTTIKKANINGRVLAQCNDILKXEM 1203
DB 1206 APGPVLLNSLNDVACEKQLQIEGLDQSMLOPYCTTIKKANINGRVLAQCNDILKXEM 1265
QY 1204 NMNFGDWHLFRSTVLEMRNASHVWPDPRFLSESSGAPGEPARRASHNELPHTELS 1263
DB 1266 NMNFGDWHLFRSTVLEMRNASHVWPDPRFLSESSGAPGEPARRASHNELPHTELS 1325
QY 1264 SQTPTYLNFPELNTLGLDEGAPRHSNLSWQSQTRRTTSLSSLSNQDSSIEISKLTDKV 1323

DB 1326 SQTPTYLNFPELNTLGLDEGAPRHSNLSWQSQTRRTTSLSSLSNQDSSIEISKLTDKV 1385
QY 1324 QAEYRDAYREYIAQMSQLEGPGSTTISGRSSPHSTYYMGQSSSGSIHNSLFEQEKGD 1383
DB 1386 QAEYRDAYREYIAQMSQLEGPGSTTISGRSSPHSTYYMGQSSSGSIHNSLFEQEKGD 1445
QY 1384 EKPDPDGRKSFILMKRGDVIDYSSGVSTNDASPLDPIITEDEKSDQSGKLLPGKSSER 1443
DB 1446 EKPDPDGRKSFILMKRGDVIDYSSGVSTNDASPLDPIITEDEKSDQSGKLLPGKSSER 1505
QY 1444 SSLFQDLDLKGSLRYQKLPSEDESGTSESDNTPLLDKDDKRAEGKVERVPKSPHS 1503
DB 1506 SSLFQDLDLKGSLRYQKLPSEDESGTSESDNTPLLDKDDKRAEGKVERVPKSPHS 1565
QY 1504 ASPIRTFIKAEYLSKGLDLDKDDSDSVRSSESSPHSLHNEVADDSOLEKANLIELED 1563
DB 1566 ASPIRTFIKAEYLSKGLDLDKDDSDSVRSSESSPHSLHNEVADDSOLEKANLIELED 1625
QY 1564 DSHSGKRGIPHSLSGLQDDPIIARMSICSDKSPSECSLIASSPENWPAQKAYNLNRT 1623
DB 1626 DSHSGKRGIPHSLSGLQDDPIIARMSICSDKSPSECSLIASSPENWPAQKAYNLNRT 1685
QY 1624 PSTVTILNNSAPANRANQNFDEMEGIRETSQVILRSPSSPNPTTIQENLKSMTKRSOR 1683
DB 1686 PSTVTILNNSAPANRANQNFDEMEGIRETSQVILRSPSSPNPTTIQENLKSMTKRSOR 1745
QY 1684 SSVYTRLSKDPPELHAAASSESTGFGBERESIL 1715
DB 1746 SSVYTRLSKDPPELHAAASSESTGFGBERESIL 1777

RESULT 2
Q9EQG6 PRELIMINARY; PRT; 1762 AA.
AC Q9EQG6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE KIDINS220.
GN Name=Kidins220;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20568256; PubMed=10998417; DOI=10.1074/jbc.M005261200;
RA Iglesias T., Cabrera-Poch N., Mitchell M.P., Naven T.J., Rozengurt E.,
RA Schiavo G.;
RT Identification and cloning of Kidins220, a novel neuronal substrate
of protein kinase D.;
RL J. Biol. Chem. 275:40048-40056 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Mitchell M., Schiavo G.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF239045; AAG35185.2; -.
DR HSSP; P09959; 1SW6.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; Ank.11.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK.11.
DR PROSITE; PS50088; ANK_REPEAT; 10.
DR PROSITE; PS50297; ANK_REPEAT; 1.
KW ANK repeat.
SQ SEQUENCE 1762 AA; 195714 MW; 0CB2689A571F8AB4 CRC64;

Query Match 94.1%; Score 8362.5; DB 2; Length 1762;
Best Local Similarity 91.6%; Pred. No. 0;
Matches 1616; Conservative 47; Mismatches 50; Indels 51; Gaps 4;
QY 1 MSVLSQSIVNVVEENIPALKLEKCKDVDERNECGQTPLMIAAQGNLEIVKELKN 60

```
Db 1 MSVLISQSVINYVEENTPALKALLECKVDNERECGQTPMLLAEBQGNVEIVKLELKN 60
Qy 61 GANCNLELDLNDWTALISASKEGHVHI VELLKCGVNLEHRDMGWGWTALMWACYKGRTDVW 120
Db 61 GANCNLELDLNDWTALISASKEGHIH VELLKSGASLEHRDMGWGWTALMWACYKGRTDVW 120
Qy 121 ELLLSHGANSPTVGLQYVYPIIWAAGRHADI VHLLLQNGAKVNCSDKYGTGTPPLVWAAR 180
Db 121 ELLLSHGANSPTVGL-YSVYPIIWAAGRHADI VHLLLQNGAKVNCSDKYGTGTPPLVWAAR 179
Qy 181 KGHLECVKHLAMGADVDOEGANSNTALIVAVKGYTOSVKEILKRNPNVNLTKDGNNTA 240
Db 180 KGHLECVKHLAMGADVDOEGANSNTALIVAVKGYTOSVKEILKRNPNVNLTKDGNNTA 239
Qy 241 LMIAASKEGHEIIVQDLLDAGTVYVNPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRG 300
Db 240 LMIAASKEGHEIIVQDLLDAGTVYVNPDRSGDTVLIGAVRGHVEIVRALLQKYADIDIRG 299
Qy 301 QDNKTALYAWVEKGNATVVRDILQCNPDTEICTKDGETPLIKATKMRNIEVVELLDKGA 360
Db 300 QDNKTALYAWVEKGNATVVRDILQCNPDTEICTKDGETPLIKATKMRNIEVVELLDKGA 359
Qy 361 KVSADVKKGDTPLHAIIRGRSKLAELLRNPKOGRLLYRNPKAGETPNYDCSHQKSIL 420
Db 360 KVSADVKKGDTPLHVAIRGRSRLAELLRNPKOGRLLYRNPKAGETPNYDCSHQKSIL 419
Qy 421 TOIFGARHLSPTETDGMGLGYDLYSSALADILSEPTMOPPCVGLYAOWGSGKSFLLKKL 480
Db 420 TOIFGARHLSPTETDGMGLGYDLYSSALADILSEPTMOPPCVGLYAOWGSGKSFLLKKL 479
Qy 481 EDEMKTFAGQOIETPLFQFQSWLIVFLTLCLCGGLGLLFAFTVHPNLGIAVSLFALLIYIF 540
Db 480 EDEMKTFAGQOETPLFQFQSWLIVFLTLCLCGGLGVFAFTVDTNLAIASLSFLALIYIF 539
Qy 541 FIVIFYGRRREGESNNAWVLSRLARHIGYLELLKLMFNPNPPELPPQTTKALPVRLPLF 600
Db 540 FIVIFYGRRREGESNNAWVLSRLARHIGYLELLKLMFNPNPPELPPQTTKALPVRLPLF 599
Qy 601 TDYNRLSVGGETSLAEMIATLSACEREFGLATRLPRVFKETDQCKKWKKTCCCLPS 660
Db 600 TDYNRLSVGGETSLAEMIATLSACEREFGLATRLPRVFKETDQCKKWKKTCCCLPS 659
Qy 661 FVIFLFIIGCIISGITTLLAIFRVPDPKHLTVNAVLSIASVVGFLAVLNCRTWQVLDL 720
Db 660 FVIFLFIIGCIISGITTLLAIFRVPDPKHLTVNAVLSIASVVGFLAVLNCRTWQVLDL 719
Qy 721 NSQRKRLHNAASKLHLKSEGFPMKVLKCEVELMARMMAKTIDSFTQNTQRLVVIIDGLDAC 780
Db 720 NSQRKRLHNAASKLHLKSEGFPMKVLKCEVELMARMMAKTIDSFTQNTQRLVVIIDGLDAC 779
Qy 781 BODKVLQMLDTRVRLFSKGPPIAFSPADPHIIKAINQNLNSVLDRDSNINGHDYMRNVH 840
Db 780 BODKVLQMLDTRVRLFSKGPPIAFSPADPHIIKAINQNLNSVLDRDSNINGHDYMRNVH 839
Qy 841 LPVFLNSRGLSNARKFLVTSATNGDVPSCDITGQEDADRVSQNSLGEMTKLGSKTALN 900
Db 840 LPVFLNSRGLSNARKFLVTSATNGDITCSDITGQEDTDRVSQNSLGEMTKLGSKTALN 899
Qy 901 RRDYTRRQORTTITROMSFOLTKLLVTEDMWFSDISPOTMRLLNIVSVTGRLLRANQIS 960
Db 900 RRDYTRRQORTTITROMSFOLTKLLVTEDMWFSDISPOTMRLLNIVSVTGRLLRANQIT 959
Qy 961 FNVDRLASWNLTRQWPPVTSWLLIYLBEETSGIPDQMTLKIYERISKNIPTTKDVEPLL 1020
Db 960 FNVDRLASWNLTRQWPPVTSWLLIYLBEETSGIPDQMTLKIYERISKNIPTTKDVEPLL 1019
Qy 1021 BIDGDIRNFEVFLSSRTPLVARDVKVFLPCTVNLDPKRLIIVADVRAAREQISIGGLAY 1080
Db 1020 BIDGDIRNFEVFLSSRTPLVARDVKVFLPCTVNLDPKRLIIVADVRAAREQINIGGLAY 1079
Qy 1081 PPLPLHEGPPRPPSGYQSPVSCVSTSPNGFAGVSPQPHSSYSGMTQPHFPYNR- 1139
```

```
Db 1080 PPLPLHEGPPRPPSGYQSPVSCVSTSPNGFAGVSPQPHSSYSGMTQPHFPYNR 1139
Qy 1140 -----GSG-----PAGPVPVLL 1151
Db 1140 PFAPLYLTPRYPPGSGQHLISRSVKTSLPRDQNGNGLPCDSGPNKQRAAVPATGSSLL 1199
Qy 1152 NSLAVADVCEKIKQIEGLDQSMLEPYCTTIKANKINGRVLQACNIDELKCKNMNFGDWH 1211
Db 1200 SSMVTVVCEKURQIEGLDQSMLEPYCTTIKANKINGRVLQACNIDELKCKNMNFGDWH 1259
Qy 1212 LFRSTVLEMRNAESHVVPEDPRFLSESSSGPAPGEPARRASHNELPHELTESSQTPYTLN 1271
Db 1260 LFRSMVLEMRSVESQVVPEDPRFLNENSSAPVPHGESARRSHTLPLTESSQTPYTLN 1319
Qy 1272 FSPFELNTGLDDEGAPRHSNLSWOSQTRTTPSLSSLSNSQDSSEIETSKLTDKVQAEYRDAY 1331
Db 1320 FSPFELNTGLDDEGAPRHSNLSWOSQTRTTPSLSSLSNSQDSSEIETSKLTDKVQAEYRDAY 1379
Qy 1332 REYIAQMSQLEGGPGSTTISGRSSPHSTYYMGQSSSGSIHNSLEQEKGDSEPKPDGDR 1391
Db 1380 REYIAQMSQLEGGTGSSTISGRSSPHSTYYIGQSSSGSIHSTLEQERKGEKELQEDGR 1439
Qy 1392 KSPFLMKRGDVTIDYSSSGVSTNDASPLDPIITEDEKSDQSGSKLLPGKSSSRSSLFQTDL 1451
Db 1440 KSFLMKRGDVTIDYSSSGVSTNEASPLDPIITEDEKSDQSGSKLLPGKSSSRPSLFQTDL 1499
Qy 1452 KLKSGRLYQKLPSEDESGTERSDNTPLLKDDKDKAEGKVERVPKSPHSAEPIRTFI 1511
Db 1500 KLKGGRLYQKLPSEDESGTERSDNTPLLKDDKDKAEGKVERVPKSPHSAEPIRTFI 1559
Qy 1512 KAKEYLDALLDKKSDSGVRSSESPNHSNLEHNEVADDSQLEKANLIELEDDSHSGKRG 1571
Db 1560 KAKEYLDALLDKKSDSGVRSSESPNHSNLEHNEVADDSQLEKANLIELEDDSHSGKRG 1619
Qy 1572 IPHLSGLQDPIIARMSICSEDKKSPSECSLIASPEENWPAQKAYNLNTPSTVTLLN 1631
Db 1620 MPHLSGLQDPIIARMSICSEDKKSPSECSLIASPEESWPAQKAYNLNTPSTVTLLN 1679
Qy 1632 NSAFANRANQNFDMEGIRETSQVILPSSSPNPTIQNELKSMTHKRSQSSSYTRLSK 1691
Db 1680 NTATNANQNFDMEGIRETSQVILPSSSPNPTIQNELKSMTHKRSQSSSYTRLSK 1739
Qy 1692 DPPELHAAASSESTGFGEERESIL 1715
Db 1740 DASELH-AASSESTGFGEERESIL 1762

RESULT 3
Q9ERD4
ID Q9ERD4 PRELIMINARY; PRT; 1715 AA.
AC Q9ERD4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Ankyrin repeat-rich membrane-spanning protein.
GN Name=ARMS;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=20585245; PubMed=11150334;
RA Kong H., Boulter J., Weber J.L., Lai C., Chao M.V.;
RT "An evolutionarily conserved transmembrane protein that is a novel
RL downstream target of neurotrophin and ephrin receptors."
RL J. Neurosci. 21:176-185(2001).
DR EMBL; AF313464; AAC34167.1; -.
DR HSSP; P03959; ISW6.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; Ank; 11.
DR PRINTS; PR01415; ANKYRIN.
```


"Prediction of the coding sequences of mouse homologues of KIAA gene:
 RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
 cDNAs identified by screening of terminal sequences of cDNA clones
 RT randomly sampled from size-fractionated libraries."
 RL DNA Res. 10:35-48(2003).
 DR EMBL; AK122478; BAC65760.1; --
 DR HSP; P09959; 1SW6.
 DR MGD; MGI:1924730; C330002I19Rik.
 DR GO; GO:0016021; C:integral to membrane; TAS.
 DR InterPro; IPR002110; ANK.
 DR Pfam; PF00023; Ank; 10.
 DR PRINTS; PR01415; ANKYRIN.
 DR SMART; SM00248; ANK; 9.
 DR PROSITE; PS50088; ANK_REPEAT; 8.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 KW ANK repeat.
 FT NON TER
 SQ SEQUENCE 1 188253 MW; 599DE2A3419D4C9D CRC64;

Query Match 87.5%; Score 7771; DB 2; Length 1693;
 Best Local Similarity 87.9%; Pred. No. 0;
 Matches 1506; Conservative 54; Mismatches 53; Indels 100; Gaps 6;

```

QY 83 HVHVEELLKCGVNLHEDMGWTALWACYKGRDVVELLSHGANSPTGLQYSYPI 142
DB 1 HIHVEELLKCGANLEHEDMGWTALWACYKGRDVVELLSHGANSPTGL-QYSYPI 59

143 IWAAGRGHADIVHLLQNGAKVNSDKYGTTPLVWAARKGHLECVKHLANGADVDEGA 202
DB 60 IWAAGRGHADIVHLLQNGAKVNSDKYGTTPLVWAARKGHLECVKHLANGADVDEGA 119

203 NSMTALIVAVKGGYTSQVKEILKNPNVNLTKDGNALMTASKEGHEIIVQDLLDAGTY 262
DB 120 NSMTALIVAVKGGYTSQVKEILKNPNVNLTKDGNALMTASKEGHEIIVQDLLDAGTY 179

263 VNIPDRSGDTVLIGAVRGCHVEIVRALLQKYADIDIRGQDNKTALYNAVEKGNATWVDI 322
DB 180 VNIPDRSGDTVLIGAVRGCHVEIVRALLQKYADIDIRGQDNKTALYNAVEKGNATWVDI 239

323 LQCPNPDTEICWKDEGTPLIKATKRNIEVBELLDKGAKVADVKKGDTPLHIAIRGRSR 382
DB 240 LQCPNPDTEICWKDEGTPLIKATKRNIEVBELLDKGAKVADVKKGDTPLHIAIRGRSR 299

383 KLABELLRNPKDGLLYRNPKAGETPNIDCSHQKSLITQIFGARHLSPETDGMGLGYD 442
DB 300 RLABELLRNPKDGLLYRNPKAGETPNIDCSHQKSLITQIFGARHLSPETDGMGLGYD 359

443 LYSSALADILSEPTMQPPICVGLYAQWGSFKLLKLEDEMKTFAQQIIEPLFQFSLWI 502
DB 360 LYSSALADILSEPTMQPPICVGLYAQWGSFKLLKLEDEMKTFAQQIIEPLFQFSLWI 419

503 VFLTLLCGGLLFAFTVHFNGLAVSLFLALLYIPFIVYFGRRREGESNNWAVLS 562
DB 420 VFLTLLCGGLLFAFTVDTNLALAVSLFLALLYIPFIVYFGRRREGESNNWAVLS 479

563 TRLARHIGYELLKLMFVNPPELPEQTTKALPVRFPLTDYNRLLSVGGETSLEMIATL 622
DB 480 TRLARHIGYELLKLMFVNPPELPEQTTKALPVRFPLTDYNRLLSVGGETSLEMIATL 536

623 SDACERBFGFLATRLFRVFKTEDTGKKKWKTCCLPSFVIFLFIIGISGITLLAIFR 682
DB 537 -----FEYRLRLKILRY-----KRNKGKTCCLPSFVIFLFIIGISGITLLAIFR 581

683 VDPKHLTVNAVLIISTASVGLAFVLCNRTWQVLDLNSQKRLHNAASKHLKKEGF 742
DB 582 VDPKHLTVNAVLIISTASVGLAFVLCNRTWQVLDLNSQKRLHNAASKHLKKEGF 641

743 MKVLKCEVELMARMAKTIDSTQNTRLVLIIDGLDACEQDKVLQMLDTRVLFSGKGPFI 802
DB 642 MKVLKCEVELMARMAKTIDSTQNTRLVLIIDGLDACEQDKVLQMLDTRVLFSGKGPFI 701

803 AIFASDPHIITKAINQNLSVLRDSNINGDHMYRNIHVLPVFLNSRGLSNARKFLVTSAT 862

```

```

Db 702 AIFASDPHIITKAINQNLSVLRDSNINGDHMYRNIHVLPVFLNSRGLSNARKFLVTSAT 761
QY 863 NGDVPFCDTTGIGEDADRRVQNSLIGEMTKLGSKTALNRDITYRRRQORITITQMSFDL 922
Db 762 NGDISCEATGVQEDADRRVQNSLIGEMTKLGSKTALNRDITYRRRQORITITQMSFDL 821
QY 923 TKLLVTEDEWFSDISPQWTRRLNIVSVTGRLLRANQISFNWDRLASINLWTEQWYRPSW 982
Db 822 TKLLVTEDEWFSDISPQWTRRLNIVSVTGRLLRANQITFNWDRLASINLWTEQWYRPSW 881
QY 983 LILVLEETEGIPDQMTLKIYERISKNIPPTKQVEPLEIDGDRNFEVLSRTPVLVA 1042
Db 882 LILVLEETEGIPDQMTLKIYERISKNIPPTKQVEPLEIDGDRNFEVLSRTPVLVA 941
QY 1043 RDVKVFLPCTVNLDPKLEIIRIADVRAAREQISIGLAYPPLPLHEGPPRAPSQVPSV 1102
Db 942 RDVKVFLPCTVNLDPKLEIIRIADVRAAREQINIGLAYPPLPLHEAPRPPSGVYQPASV 1001
QY 1103 CSST-SFNGPPAGGVSPQPHSSYISGWTGQHPFYNR----- 1139
Db 1002 CSSSASFGNPPGCVSPQPHSSYISGLSGQHPFYNRPPFAPVLYTPRYYPGSGSHLIS 1061
QY 1140 -----CSG 1142
Db 1062 RSVKTSPLRDQNNGLVIEDAAEGLPSPASPASREKSWTKQLMELCDSGFNKQRQASV 1121
QY 1143 PAPGVVLLNSLVDAVCEKLRQIEGLDQSMPLPOYCTTIKANINGRVLACNIDELKKE 1202
Db 1122 PATGTSLLSSMTVDVCEKLRQIEGLDQSMPLPOYCTTIKANINGRVLACNIDELKKE 1181
QY 1203 MNMFGDWHLFRSTVLEMRNABSHVPEDPFRFLSESSGPPHGEPPARRASHNELPHTL 1262
Db 1182 MAMNFGDWHLFRSMVLEMRNABSHVPEDPFRFLSESSGPPHGEPPARRASHNELPHTL 1241
QY 1263 SSQTPYTLNPSFEELNTGLDEGAPRHSNLSWQOTRTPSLSSINQDSSIEISKLTDK 1322
Db 1242 SSQTPYTLNPSFEELNTGLDEGAPRHSNLSWQOTRTPSLSSINQDSSIEISKLTDK 1301
QY 1323 VQAEYRDAYREYIAQMSQLEGPGSTTISGRSSPHSTVYMGQSSSGGSIHNLQEKGD 1382
Db 1302 VQAEYRDAYREYIAQMSQLEGPGSTTISGRSSPHSTVYMGQSSSGGSIHNLQEKGD 1361
QY 1383 SEPKDDGRKSFMLMKRGDVIDYSSSGVSTNDASPLDPITEDEKSDQSGSKLLPKKSS 1442
Db 1362 SELKQEGGRKSFMLMKRGDVIDYSSSGVSTNEASPLDPITEDEKSDQSGSKLLPKKSS 1421
QY 1443 RSSLFQTDLLKSGSLRYQKLPSEDESGTSEESDNTPLDKDKRKAEGKVERVPKSPH 1502
Db 1422 RPSLFQTDLLKSGSLRYQKLPSEDESGTSEESDNTPLDKDKRKAEGKVERVPKSPH 1481
QY 1503 SAEPIRTFKAKEYLSDALLDKDSDSGVRSSESSPNHSLHNEVADDQLEKANLIELE 1562
Db 1482 SVEPIRTFKAKEYLSDALLDKDSDSGVRSSESSPNHSLHNEVADDQLEKANLIELE 1541
QY 1563 DDHSGRGKIPHLSGLQDPIIARMSICSDKSPSCSLIASSPEENWPACQAYNLNR 1622
Db 1542 DEHSGRGKIPHLSGLQDPIIARMSICSDKSPSCSLIASSPEENWPACQAYNLNR 1601
QY 1623 TPSITVTLNNSAPANRANQFDEMEGIRETSQVILRFPSSSPNPTTIQENLKSMTWKRSQ 1682
Db 1602 TPTSTVTLNNTAPTNRANQFDEIEGIRETSQVILRFPSSSPNPTTIQENLKSMTWKRSQ 1661
QY 1683 RSSYTRLSKDPPELHAAASSTGFGGERESIL 1715
Db 1662 RSSYTRLSKDPPELHAAASSTGFGGERESIL 1693

RESULT 5
Q80SX9
ID Q80SX9 PRELIMINARY; PRT; 1554 AA.
AC Q80SX9
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

```



```
Db 1261 TEEDEKSDQSGSKLLPGKSSERPSPSLFQTDLKLKSGGLRYOKLPSEDESGTEESDNTPL 1320
Qy 1481 LKODKORKAEGKVERVPKSPHSAEPRTFKAKEYLSDALLDKKSDSDSGVRSSESPN 1540
Db 1321 LKDDKDKAEGKAERVAKSPHSAEPRTFKAKEYLSDALLDKKSDSDSGVRSSESPN 1380
Qy 1541 HSLNEVADDDQLEKANLIELEDDSHSGKRGIPHSLSGLQDPPIARMSCSEDKKSPSEC 1600
Db 1381 HSLNEAADDDQLEKANLIELEDDSHSGKRGMPHPSLSGLQDPPIARMSCSEDKKSPSEC 1440
Qy 1601 SLIASSPENPACOKAYNLARTSTVTNLNNSAPANRANONFEDMEGIRTSQVILRPS 1660
Db 1441 SLIASSPENPACOKAYNLARTSTVTNLNNSAPANRANONFEDMEGIRTSQVILRPS 1500
Qy 1661 SSPNPTIQENLKMTHKRSQSSYTRLSKDPPELHAAASSESTGFGEERESIL 1715
Db 1501 PSPNPTAQVENLKMTHKRSQSSYTRLSKDASELH-AASDSTGFGEERESIL 1554

RESULT 6
Q7T163 ID Q7T163 PRELIMINARY; PRT; 1680 AA.
AC Q7T163;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE SI:d2119J18.2 (Novel protein similar to rat kinase D-interacting
DE substance of 220 kDa (KIDINS220) ).
GN Name:SI:d2119J18.2;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Babbage A.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL954721; CAE17588.1; -.
DR HSSP; P42773; 1IHB.
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00023; Ank; 11.
DR SMART; SM00248; ANK; 11.
DR PROSITE; PS50088; ANK_REPEAT; 10.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; UNKNOWN_1.
KW ANK repeat; Kinase.
SQ SEQUENCE 1680 AA; 185948 MW; 4F88AB3377DD5E60 CRC64;

Query Match 70.8%; Score 6289.5; DB 2; Length 1680;
Best Local Similarity 72.5%; Pred. No. 2.6e-239;
Matches 1258; Conservative 166; Mismatches 229; Indels 81; Gaps 27;

Qy 1 MSVLISQVINYVEENIPALKALLEKCKVDNERNECQOTPLMIAAEOGNLEIVKELIKN 60
Db 9 MTLIAIQNLFSYVEENLAAVKVHLDFKEVDGSDNGQTFPLMLASQGSLEIVQELIRR 68
Qy 61 GANCNLEDLDNWTALISASKEGHVHIVVELLKCGVNLHHRDMGWTALMWCYKGRDYY 120
Db 69 GANYNLDDVDCWSALISAAKEGHVEVVKELLENSAYIEHRDMGWTALTWASYKGRVEVA 128
Qy 121 ELLLSHGANSVTVGLQYVYPIIIWAAGRGHADIIVHLLILONGCAKNCSDKYGTTFPLWAAR 180
Db 129 TVLLENGANPNNTQQQYVYVPIIIWAAGRGHAEIVKLEHGAKNCSDKYGTTFPLWAAR 188
Qy 181 KGHLECVKHLAMGADVDQEGANSMTALIVAVKGGYTSQVKEILKRNPNVNLTKDGNTA 240
Db 189 KGHVDCWHLLENGADVQEGANSMTALIVAVKGGYTEVVKELKRNPNVNLTKDGNTA 248
Qy 241 LMTIASKEGTEIVODLLDAGTYVNIIPDRSGDTVLIGAVRGGHVEIVRALLQKYADIDIRG 300
Db 1366 EICKLTDKQAEYRNAYEDYIASMSQELG-----MEKPPVPPFVLSQMLHSSS----- 1366
```

```
Db 249 LMTIAAEGYTEIVODLLDAGTYVNIIPDRSGDTVLIGAVRGGHVEIVRALLHKKYADIDIRG 308
Qy 301 QDNKTALYWAKEGNATWVRDILQCNPDTEICTKGETPLIKATQRNIEVVVELLDKGA 360
Db 309 QENKTALYWAKEGNATWVRDILQCNPDTEICTKGETPLIKATQRNIEVVVELLDKGA 368
Qy 361 KVSADVKKGDPTPLHAIAGRSRKLAEALLRNPKDGRLLYRPNKAGETPNYIDCSHOKSIL 420
Db 369 KVSADVKKGDPTPLHAIAGRSRKLAEALLRNPKDGRLLYRPNKAGETPNYIDCSHOKSIL 428
Qy 421 TQIFGARHLSSTEDGMLGYDLYSSALADILSEPTMQPPICVGLYAGWGSKSLKKL 480
Db 429 TQIFGARHLSSTEDGMLGYDLYSSALADILSEPTMQPPICVGLYAGWGSKSLKKL 488
Qy 481 EDEMKTFAGQOIEPLFQFESWLI VFLTLLCGGLGLLPFAFTVHPNLGIAVLSIFLALLYF 540
Db 489 EDEMKTFAGQOIEPLFQFESWLI VFLTLLCGGLGLLPFAFTVHPNLGIAVLSIFLALLYF 548
Qy 541 FIVYFGRREGESWNNAWVLSLRLARHIGYLELLKLMFVNPPELPBQTTKALPVRFLF 600
Db 549 FVYVYFGRREGESWNNAWVLSLRLARHIGYLELLKLMFVNPPELPBQTTKALPVRFLF 608
Qy 601 TDYNRLSSVGGETSLAEWIAIATLSDACEREGFLATRLFRVFKTEDTQKKKWKTCCLPS 660
Db 609 TDYNRLSSVGGETSLAEWIAIATLSDACEREGFLATRLFRVFKTEDTQKKKWKTCCLPS 668
Qy 661 FVIFLFIIGCTIIGITLAIIPRVDPKHLTVNAVLSIASVVGGLAVFLNCRTWQVLDL 720
Db 669 FVIFLFIIGCTIIGITLAIIPRVDPKHLTVNAVLSIASVVGGLAVFLNCRTWQVLDL 728
Qy 721 NSQKRLHNAASKLHLKSEGMKVLCEVELMARMARKTIDSFQNTQNTLRLVVIDGLDAC 780
Db 729 NSQKRLHNAASKLHLKSEGMKVLCEVELMARMARKTIDSFQNTQNTLRLVVIDGLDAC 788
Qy 781 EQDKVQLMDTVRVLPKSGPPIAFASDPHIIKAINONLNSVLRDSNINGHDYMRNVH 840
Db 789 EQDKVQLMDTVRVLPKSGPPIAFASDPHIIKAINONLNSVLRDSNINGHDYMRNVH 848
Qy 841 LPVFLNSRGLSNARKFLVTSATNDGVPSCDTGQEDADRDVRSNLSGEMTKLASKTALN 900
Db 849 LPVFLNSRGLSNARKFLVTSATNDGVPSCDTGQEDADRDVRSNLSGEMTKLASKTALN 906
Qy 901 RRDYRERQMQRTITRQMSFDLTLLVTEWDFSDISPTQWRLRLNIVSVTGRLLRANQIS 960
Db 907 RRDYRERQMQRTITRQMSFDLTLLVTEWDFSDISPTQWRLRLNIVSVTGRLLRANQIS 966
Qy 961 FNDRLASWINLTQWNPYRSLWILYLEETEGIPDQMTLKIYERISKNIPTTKDVEPLL 1020
Db 967 FNDRLASWINLTQWNPYRSLWILYLEETEGIPDQMTLKIYERISKNIPTTKDVEPLL 1026
Qy 1021 EIDGDIENFEVFLSSRTPVLVARDVKVLPCTVNLDPKLEI IADVRAAREQISIGLAY 1080
Db 1027 EIDGDIENFEVFLSSRTPVLVARDVKVLPCTVNLDPKLEI IADVRAAREQISIGLAY 1086
Qy 1081 PPLPLHGPFPRAFGSYQPPSVCSST-SFNGPFAAGVVPVSPHSSYSGMTGPQHPFYNR 1139
Db 1087 PPLPLHGPFPRAFGSYQPPSVCSST-SFNGPFAAGVVPVSPHSSYSGMTGPQHPFYNR 1144
Qy 1140 GSGPAPG-PVVLNLSNLDVAVCEKLEKQIBGLDQSMPLPYQYCTTIKKANINGRVLAQCNI 1197
Db 1145 GSASVSGTSPILSSMSTDVCERVKLIDGIDQNLISQYATATIKKANINGRVLSQCNI 1204
Qy 1198 ELKKEMMNFGDWHLFSTVLEMRNAESHVVPDPFLSESSCPAPHEGPAR---ASH 1254
Db 1205 ELKKEMMNFGDWHLFSTVLEMRNAESHVVPDPFLSESSCPAPHEGPAR---ASH 1262
Qy 1255 NELPTELSSQTPYTLNFSPEELNTLGLDEGAPHSNLSWOSQTRRTPSLSSLSQDSSI 1314
Db 1263 GVAGNTDTPM--YFNFLNFSPEELNSVGLER-PPRHVNATWMTGTHRTPSMSSLSQSSN 1319
Qy 1315 EISKLTQKQAEYRDAYRIYAQMSQLEGGPGSTTISGRSPSHSYTYMGSSSGGSIHNS 1374
Db 1320 EICKLTDKQAEYRNAYEDYIASMSQELG-----MEKPPVPPFVLSQMLHSSS----- 1366
```

```
QY 1375 LEOEKKQSEPKDDCKRGSFLMKRG-----DVIDYSSSGVSTNDASPLDPITEDEKSP 1428
Db 1367 --EDKKKGNDQ--DGRKS--VSRGSGTSGSDNTDYASA-----DAATLDPIITEDEKVD 1416
QY 1429 QSGSKLLPGKSS--ERSSLFQ--TDLKL--GSLGRYQKLPSDESGTSESDNTPLKDDK 1485
Db 1417 HGSKSLGRKTSQDKVSLFQADLKLKAGGSGRYQKLTSDDES--ESDNPALDKGK 1474
QY 1486 DRKAEQKVERPKSPHSABPIRTFIKAKEYLSDALLDKOSSDSGVRSSSESSPNHSLN 1545
Db 1475 --KPEAKA-----SDGDRSLTKGDKYLS---DKOSSDSGVRSSSESSPNHSLQD 1519
QY 1546 EVADDSQLEKANLIEEDSHSGKRGIPHSLSGLQDPIIARMSICSDKSPSECSLIAS 1605
Db 1520 EADLSQSERANLIEDEENSARKRGLPNSLSGLQDPTIARMSICSD-----QCSLLAS 1574
QY 1606 SPENNPACOKAYNLARTPTVTLLN--NSAPANRANQNFDEMEGIRET--SQVILRPSS 1661
Db 1575 SPESWPS--SKSYNLARTPTNTLNNTNAQGNHRIQPSDSNTTSTTSGSVIINPGT 1633
QY 1662 SPNPTTIONENLKMTHKRSQSSYTRLSKOPPELHAASSESTGFGERESIL 1715
Db 1634 STTSATTQENVRVHLKRLNGP-----DPPEI-LKVSSTVTFGERESIL 1680

RESULT 7
Q6P7Y1
ID Q6P7Y1 PRELIMINARY; PRT; 1672 AA.
AC Q6P7Y1;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Hypothetical protein zgc:63531.
GN ORFNames=zgc:63531;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Whole body;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klauener R.D., Collins F.S., Wagner L., Schmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Whole body;
RA Strausberg R.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC061450; AAH61450.1;
DR HSSP; Q60773; 1AP7.
DR ZFIN; ZDB-GENE-030131-7824; zgc:63531.
DR InterPro; IPR002110; ANK.
```

```
DR InterPro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00023; Ank; 10.
DR PRINTS; PR01415; ANKRN.
DR SMART; SM00248; ANK; 11.
DR PROSITE; PS0088; ANK REPEAT; 10.
DR PROSITE; PS00297; ANK REP REGION; 1.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; UNKNOWN_1.
KW ANK repeat; Hypothetical protein.
SQ SEQUENCE 1672 AA; 185035 MW; 75BECF10280B410B CRC64;

Query Match 70.6%; Score 6275.5; DB 2; Length 1672;
Best Local Similarity 72.4%; Pred. No. 1.2e-298;
Matches 1256; Conservative 166; Mismatches 231; Indels 81; Gaps 27;

QY 1 MSVLISQSVINVEENIPALKALEKCKVDVERNECGGTPLMIAAEOGNLEIVKELIKN 60
Db 1 MTTLATQNFVSYVEENLAALKVHLDPKFEVDGRSDNGQTPLMLASEQGSLEIVQELIRR 60
QY 61 GANCNLEDNDNTALISASKEGHVHIVELLKCGVNLEHRDMGGMTALMWACVKGRTDVV 120
Db 61 GANVNDDVDCCSALLISAKEGHEVVEVKELLENSAIEHRDMGGMTALTWASYKGRVEVA 120
QY 121 ELLSHGANPSVTGLQYSVYPIIWAAGRGHADIHLLQNGAKVNCSDKYGTTPPLVWAAR 180
Db 121 TVLLEANGNPNTTGGQYSVYPIIWAAGRGHAEIVKLLBHGAKVNCSDKYGTTPPLVWAAR 180
QY 181 KGHLECVKHLAMGADVDOEGANSMTALIVAKGGYTQSVKEILKKNPNVNLTKDGNTA 240
Db 181 KGHYDCVHLLENGADVDOEGANSMTALIVAKGGYTVVVKELKKNPNVNLTKDGNTA 240
QY 241 LMTASKEGHEIIVQDLDLAGTYVNIIPDRSGDTVLICAVRGHVEIVRALLQKADIDIRG 300
Db 241 LMTAAKEGTEIIVQDLDLAGTYVNIIPDRSGDTVLICAVRGHVEIVRALLKHADIDIRG 300
QY 301 QNKATLYWAVEKGNATWVRDILQCNPDTEICTKOGTEPLIKATKMRNIEVVELLDKGA 360
Db 301 QENKATLYWAVEKGNATWVRDILQCNPDTEITTKDSETPLIKATKMRNIEVVELLDKGA 360
QY 361 KVSADVKKGDTPLHIAIRGSRKLABLLRNPKDGLLRPNKAGETPNYIDCSHOKSIL 420
Db 361 KVSADVKKGDTPLHIAIRGSRRLAELLRNPKDGLLRPNKAGETPNYIDCSHOKSIL 420
QY 421 TQIFGARHLSPTETDGMGLDYLLSALADILSEPTMQPPICVGLYAQWGSGRSFLKKL 480
Db 421 TQIFGARHLSPTESDGMGLDYLLSALADILSEPTMQPPICVGLNTQWGSGRSFLKKL 480
QY 481 EDEMTKFAQQIEPLPQFSWLIVFLTLCCGGLGLLFAFTVHPNLGIAVSLSPALLIYIF 540
Db 481 EDEMTKFAQQIEPLPQFSWLIVLLSLLCGSVALVGLFTVDPKLAIAISLSLALLIYVF 540
QY 541 FVIVYFGRREGESWNWVLSLRLARHIGYLELLIKLMFVNPPPELPEQTTKALPVRFLE 600
Db 541 FVIVYFGRREGESWNWVLSLRLARHIGYLELLIKLMFVNPPPELPEQTTKALPVRFLE 600
QY 601 TDYNRSLSSVGETSLAEMIATISDACERFGLATRLFRVFKTEDTQGGKKWKTKCLPS 660
Db 601 TDYNRSLSSVGETSLAEMIATISDACERFGLATRLFRVFKTEDTQGGKKWKTKCLPS 660
QY 661 FVIFLFIIGCIISGITLLAIFRVDPKHLTVNAVLISIASVVGFLAVLNCRTWQVLDL 720
Db 661 FVIFLFIIGCLINGMALLAVFKVQGNQVNAVLVSMASVVGGLALLNCRTWQVLDL 720
QY 721 NSQKRLHNAASKLHKLKSEGPKVULKCEVELMARMAKIDISTQNTQRLVLIIDGLDAC 780
Db 721 NSQKRLHNAASKLHKLKSEGPKVULKCEVELMARMAKIDISTQNTQRLVLIIDGLDAC 780
QY 781 EODKVLQMLDTRVLFSGKPFPTAI PASDPHIIKAINQNLNSVLRDSNTHGYDMRNIVH 840
Db 781 EQDKVLQMLDTRVLFSGKPFPTAI PASDPHIIKAINQNLNSVLRDSNTHGYDMRNIVH 840
QY 841 LPVFLNSRGLSNARKFLVTSATNGDVPDCSDTTGIGQEDADRRVSQNSLGEWTKLGSKTALN 900
Db 841 LPVFLNSRGLSNAKKWACAPANGET--GNSEGWHEELDRKLSQNSLIGDQTKFGSKTTLN 898
```

QY 901 RDTYRRQMQRTTIQMSFOLTKLLVTEWFDSDISQPMRRLNIVSVTGRLLRANQIS 960
DB 899 RDTYRRQMQRSVTRQMSFOLTKLLVTEWFDSDISQPMRRLNIVSVTGRLLRANQIS 958
QY 961 FNDWDLASWILNTEQMPYRTSWLILYLBETEGIPQMTLKIYIERISKNIPTTKDVEPLL 1020
DB 959 FNDWDLASWILNTEQMPYRTSWLILYLBETEGIPQMTLKIYIERISKNIPTTKDVEPLL 1018
QY 1021 BIDGIRNFVFLSRTPLVARDVKVFLPCTVNDPKLRELIADVRAAREQISIGGLAY 1080
DB 1019 EIDGVRSEFVFLSRTPLVARDIRTFPCTVNDPKLRELIADVRAAREQVNWAGTY 1078
QY 1081 PPLPLHEGPPRAGSYGPPSVCSST-SFNGPPAGGVVSPPHSSYYSGMTGPQHPFFYR 1139
DB 1079 PTLPLQEGRP--ISMYSQSSACSPASFGNGPYNPVSPHSAFSGMAGPQHPFFYR 1136
QY 1140 GSGPAPG--PVVLLNSLVDAVEKLEKQIEGLDQSMPLPOYCTTIKANINGRVLACQND 1197
DB 1137 GSASVVSCTPISLLSSMTDVCERVKLIDGIDQNLISQYATTIKANINGRVLSCQND 1196
QY 1198 ELKEMNNFQDWHLFRSTVLEMRNAESHVVPEDPRFLSESSSGPAPGEPARR--ASH 1254
DB 1197 ELKEMNNFQDWHLFRSTVLEMRNAESHVVPEDPRFLSESSSGPAPGEPARR--ASH 1254
QY 1255 NELPHELSSQTPYTLNFSFELNLTGLDEGAPRHSNLSWOSQTRRTPLSLSLNSQDSSI 1314
DB 1255 GVAGNTDTPM--YFNLSPELSNVGLEE-PPRHVNATWGCATHTRTPSMSSLSQESSN 1311
QY 1315 EISKLTQVQBYRDAYREYIAQMSQLEGGPGSTTISGRSPHSTYTYMGQSSSGSIHNS 1374
DB 1312 EICKLTQVQBYRDAYREYIAQMSQLEGGPGSTTISGRSPHSTYTYMGQSSSGSIHNS 1374
QY 1375 LEQKGDSEKPKDDGRKSLFKMGK-----DVIDYSSSGYSTNDASPLDPTIREDKSD 1428
DB 1359 --EUKKQDNGDQ--DGRKS-VSKRGSTKSGSDNTDYASA-----DAATLDPITBEDEKVD 1408
QY 1429 QSGSKLLPGKISS-ERSLFO--TDLKLG-SGLRYQKLPSEDESGETEESDNTPLKDDK 1485
DB 1409 HGSSKSLGRKTSQDKVSLFQADLKLKAGGSRVQKLTSDDES--EESDNPALLKDGK 1466
QY 1486 DRKAGKVERVPKPEHSAEPIRTFIKAKYLSDALDKDSDSGSVRSSESSPNHSLN 1545
DB 1467 --KPEAKA-----SDGGDRSLTKGKVL-----DKDSDSGSVRSSESSPNHSLQD 1511
QY 1546 EVADDSOLEKANLELDDSHSGKRGIPHSLSGLQDPIIARMSICSEDKKSPSECSLIAS 1605
DB 1512 BEADLSOSERANLELDENSARKRGLPNSLSGLQDPTIARMSICSE-----QCSLIAS 1566
QY 1606 SPEENWPACQAKYNLNRTPTSTWTLNN--NSAPANRANQNFDEMEGIRET--SQVILRPSS 1661
DB 1567 SPESWPS-SKSYNLNRTPTSTWTLNNNTNNAQGNHIRQPSDSSNTTSTTSGDVIINPGT 1625
QY 1662 SPNPTTIONELKSWTHKRSORSSYTRLSKDPPELHAAASSESTGPGERESIL 1715
DB 1626 STTSATTQENVRVVLKRLGNPG-----DPPEI-LKVSSETVTFGERESIL 1672

RESULT 8

Q9UF42 ID Q9UF42 PRELIMINARY; PRT; 1031 AA.
AC Q9UF42; DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Hypothetical protein DRFZp434F0621.
GN Name=DRFZp434F0621;
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

RC TISSUE-Testis;
RA Blum H., Baurersachs S., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL133620; CAB63746.1; --
DR PIR; T43458; T43458.
DR HSSP; P09959; ISW6.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; Ank; 11.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 11.
DR PROSITE; PS00088; ANK_REPEAT; 10.
DR PROSITE; PS02097; ANK_REPEAT_REGION; 1.
KW ANK repeat; Hypothetical protein.
SQ SEQUENCE 1031 AA; 115310 MW; D98866461C13A2F5 CRC64;

Query Match 58.6%; Score 5205.5; DB 2; Length 1031;
Best Local Similarity 98.5%; Pred. No. 1.6e-246;
Matches 1008; Conservative 2; Mismatches 0; Indels 13; Gaps 1;

QY 1 MSVLISQSVINYVEENIPALKALLEKCKVDNERNECCQTPLMIAAEEQGNLEIVKELIKN 60
DB 1 MSVLISQSVINYVEENIPALKALLEKCKVDNERNECCQTPLMIAAEEQGNLEIVKELIKN 60
QY 61 GANCNLELDNWTALISASKEGHVHIVEELKCGVNLKCGVNLKCGVNLKCGVNLKCGVNL 120
DB 61 GANCNLELDNWTALISASKEGHVHIVEELKCGVNLKCGVNLKCGVNLKCGVNLKCGVNL 120
QY 121 ELLLSHGANSVTVGLQSVYPIIWAAGRGHADIHLLLLONGAKVNCSDKYTTPLVWAAR 180
DB 121 ELLLSHGANSVTVGLQSVYPIIWAAGRGHADIHLLLLONGAKVNCSDKYTTPLVWAAR 180
QY 181 KGHLECVKHLAMGADVDQEGANSWTALIVAVKGYTQSVKEILKRNPNVNLTKDGNNTA 240
DB 181 KGHLECVKHLAMGADVDQEGANSWTALIVAVKGYTQSVKEILKRNPNVNLTKDGNNTA 240
QY 241 LMTASKEGTEIIVODLLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQYADIDIRG 300
DB 241 LMTASKEGTEIIVODLLDAGTYVNIIPDRSGDTVLIGAVRGHVEIVRALLQYADIDIRG 300
QY 301 QDNKTALYNAVEKGNATVRDILQCNPDTEICTKGDTPLIKATMRNIEVVELLLDKGA 360
DB 301 QDNKTALYNAVEKGNATVRDILQCNPDTEICTKGDTPLIKATMRNIEVVELLLDKGA 360
QY 361 KVSADVKKGTPLHIAIRGRSKLAELLLRNPKDGRLLYRNKAGETPYNIDCSHOKSIL 420
DB 361 KVSADVKKGTPLHIAIRGRSKLAELLLRNPKDGRLLYRNKAGETPYNIDCSHOKSIL 420
QY 421 TQIFGARHLSPTETDGMGLYDLYSSALADILSEPTWQPPICVGLYAGWGSFKSFLKKL 480
DB 421 TQIFGARHLSPTETDGMGLYDLYSSALADILSEPTWQPPICVGLYAGWGSFKSFLKKL 480
QY 481 EDEKKTAGOOIEPLFQPSWLIIVFLTLLCGGLLPAFTVHPNLGIAVSLFLALYIF 540
DB 481 EDEKKTAGOOIEPLFQPSWLIIVFLTLLCGGLLPAFTVHPNLGIAVSLFLALYIF 540
QY 541 FVIYFGRRGESNNWAWLSTELARHIGVLELLKLMFVNPPELPEQTAKLPVRLFL 600
DB 541 FVIYFGRRGESNNWAWLSTELARHIGVLELLKLMFVNPPELPEQTAKLPVRLFL 600
QY 601 TDYNRLSSVGGETSLEAMIAATLSACEREFGLATRLFRVFKTEDTQKKKKKTKCCLPS 660
DB 601 TDYNRLSSVGGETSLEAMIAATLSACEREFGLATRLFRVFKTEDTQKKKKKTKCCLPS 660
QY 661 FVIFLFIIGCIISGITLLAIIPRVDPKHLTVNAVLIASVVGGLAFVNLNCRTWQVLSLL 720
DB 661 FVIFLFIIGCIISGITLLAIIPRVDPKHLTVNAVLIASVVGGLAFVNLNCRTWQVLSLL 720
QY 721 NSQKRLHNAASKLHLKLSGEMKVLKCEVELMARMARKTIDSFQNTQRLVVIIDGDLAC 780
DB 721 NSQKRLHNAASKLHLKLSGEMKVLKCEVELMARMARKTIDSFQNTQRLVVIIDGDLAC 780
QY 781 EQDKVQLMDTVRVLFSGKGFPIAFASDPHIIKAINQNLSVLRDSNINSHDYMNVIVH 840
DB 781 EQDKVQLMDTVRVLFSGKGFPIAFASDPHIIKAINQNLSVLRDSNINSHDYMNVIVH 840

```
Db 781 EQDKVLQMLDTRVLFSGKPFIFAFAADPHIIKAINQNLNSVLRDSNINGHDYMRNIHV 840
Qy 841 LPVFLNSRGLSNARKFLVTSATNGDVPSCSTTGIQEDADRRVSONSLGEMTKLGSXTALN 900
Db 841 LPVFLNSRGLSNARKFLVTSATNGDVPSCSTTGIQEDADRRVSONSLGEMTKLGSXTALN 900
Qy 901 RRTYRRRQWRTITRQMSFDLTKLVTEWDFSDISPTQMRRLNINVSVTGRLLRANQIS 960
Db 901 RRTYRRRQWRTITRQMSFDLTKLVTEWDFSDISPTQMRRLNINVSVTGRLLRANQIS 960
Qy 961 FNWDLASWINLQEQPYRTSWILYLBEETEGIPDQWTLKTIYER-----IS 1007
Db 961 FNWDLASWINLQEQPYRTSWILYLBEETEGIPDQWTLKTIYERCCGADSCDRDRIGIS 1020
Qy 1008 KNI 1010
Db 1021 KSV 1023

RESULT 9
Q72322 PRELIMINARY; PRT; 691 AA.
AC Q72322;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein DKFZp686H14204;
GN Name=DKFZp686H14204;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human fetal brain;
RA Lauber J., Bahr A., Mewes H.W., Weil B., Amid C., Osanger A., Fobo G.,
RA Han M., Wiemann S.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX538189; CAD98059.1; -
DR InterPro; IPR010993; SAM_homology.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 691 AA; 75760 MW; F5783F24DDEA6027 CRC64;

Query Match 40.3%; Score 3583; DB 2; Length 691;
Best Local Similarity 100.0%; Pred. No. 2.9e-167;
Matches 691; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1025 DIRNEFVLSRTPVLVARVKVFLPCTVNLDPKLRRIIADVRAAREQISIGGLAYPPLP 1084
Db 1 DIRNEFVLSRTPVLVARVKVFLPCTVNLDPKLRRIIADVRAAREQISIGGLAYPPLP 60

Qy 1085 LHGGPPRAPSGYQPPSVCSSTSFNGPFGAGVVSPPHSSYSGMTGPQHPFFNRRSGGA 1144
Db 61 LHGGPPRAPSGYQPPSVCSSTSFNGPFGAGVVSPPHSSYSGMTGPQHPFFNRRSGGA 120

Qy 1145 PGPVLLNSLVDAVCEKQLKQIEGLDQSMPLPYCTTIKKANINGRVLACNIDELKEMN 1204
Db 121 PGPVLLNSLVDAVCEKQLKQIEGLDQSMPLPYCTTIKKANINGRVLACNIDELKEMN 180

Qy 1205 MNFGDWHLFSTVLEMRNAESHVVPDPFLSSSSGAPAGHPAPARRASHNELPHTLELS 1264
Db 181 MNFGDWHLFSTVLEMRNAESHVVPDPFLSSSSGAPAGHPAPARRASHNELPHTLELS 240

Qy 1265 QTPVTNLNFSPEELNTGLDGAAPHNSLWSQSTRTTPSLSSLSNDSQSSIEISKLTDKVQ 1324
Db 241 QTPVTNLNFSPEELNTGLDGAAPHNSLWSQSTRTTPSLSSLSNDSQSSIEISKLTDKVQ 300

Qy 1325 AEYRDAYREYIAQMSQLEGPGGTTISGRSSPHSTHYMGSSGGSIHNLGEQKGDSE 1384
Db 301 AEYRDAYREYIAQMSQLEGPGGTTISGRSSPHSTHYMGSSGGSIHNLGEQKGDSE 360

Qy 1385 PKDDGKSLMKRGDVIDYSSSGVSNDSASPLDPIITEDEKSDQSGSKLLPGKKSERS 1444
```

```
Db 361 PREDDGKSLMKRGDVIDYSSSGVSTNDASPLDPIITEDEKSDQSGSKLLPGKKSERS 420
Qy 1445 SLFQTDLKLKSGRLYQKLPSEDESGTEESDNTPLKDDKDKAKGKVERVPKSPHSA 1504
Db 421 SLFQTDLKLKSGRLYQKLPSEDESGTEESDNTPLKDDKDKAKGKVERVPKSPHSA 480
Qy 1505 EPRTPIKAKYLSALLDKDSSDGVSRSSPNHSLHNEVADDSQLEKANLIELEDD 1564
Db 481 EPRTPIKAKYLSALLDKDSSDGVSRSSPNHSLHNEVADDSQLEKANLIELEDD 540
Qy 1565 SHSGRGKIGPHSLGLODPIIARMSICSEDKSPSECSLIASSPEENWPACQKAYNLNRT 1624
Db 541 SHSGRGKIGPHSLGLODPIIARMSICSEDKSPSECSLIASSPEENWPACQKAYNLNRT 600
Qy 1625 STVTLLNNSAPANRANQNFDEMEGIRETSQVILRPSSSNPPTIQENILKSMTHKSORS 1684
Db 601 STVTLLNNSAPANRANQNFDEMEGIRETSQVILRPSSSNPPTIQENILKSMTHKSORS 660
Qy 1685 SYTRLKSKOPPELHAAASSESTGFERESIL 1715
Db 661 SYTRLKSKOPPELHAAASSESTGFERESIL 691

RESULT 10
Q6MZU2 PRELIMINARY; PRT; 664 AA.
ID Q6MZU2;
AC Q6MZU2;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein DKFZp686A19189 (Fragment).
GN Name=DKFZp686A19189;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human esophagus tumor;
RA Bloecker H., Boecker M., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX640878; CAE45935.1; -
DR InterPro; IPR010993; SAM_homology.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 664 AA; 72879 MW; 678BA3EF27DA1BDF CRC64;

Query Match 36.1%; Score 3211; DB 2; Length 664;
Best Local Similarity 94.0%; Pred. No. 4.7e-149;
Matches 624; Conservative 0; Mismatches 2; Indels 38; Gaps 1;

Qy 1090 PRAPSGYQPPSVCSSTSFNGPFGAGVVSPPHSSYSGMTGPQHPFFYNNR----- 1139
Db 1 PRAPSGYQPPSVCSSTSFNGPFGAGVVSPPHSSYSGMTGPQHPFFYNNRPFAPLYTP 60
Qy 1140 -----GSGPAGPFWLLNSLVDAVCEKQLKQIEGLDQ 1171
Db 61 RYYPGSGHLSRPSVKTSIPLRQNNGLSGPAGPFWLLNSLVDAVCEKQLKQIEGLDQ 120
Qy 1172 SMLPYQCTTIKKANINGRVLACNIDELKEMNFGDWHLFSTVLEMRNAESHVVPDP 1231
Db 121 SMLPYQCTTIKKANINGRVLACNIDELKEMNFGDWHLFSTVLEMRNAESHVVPDP 180
Qy 1232 PRFLSSSSGAPAGHPAPARRASHNELPHTLELSQTPYTLNFSPEELNTGLDGAAPHNS 1291
Db 181 PRFLSSSSGAPAGHPAPARRASHNELPHTLELSQTPYTLNFSPEELNTGLDGAAPHNS 240
Qy 1292 LSWQSQTRTPSLSSLSNDSQSSIEISKLTDKQAEYRDAYREYIAQMSQLEGPGGTTIS 1351
Db 241 LSWQSQTRTPSLSSLSNDSQSSIEISKLTDKQAEYRDAYREYIAQMSQLEGPGGTTIS 300
```

```
QY 1352 GRSSPHSTYMGQSSSGSIHSNLEQKKGKQSEPKDGRKSFMLMKRGVDVYSSSGYST 1411
DB 301 GRSSPHSTYMGQSSSGSIHSNLEQKKGKQSEPKDGRKSFMLMKRGVDVYSSSGYST 360
QY 1412 NDASPLDPTTEDEKSDGSKLLPGKKSSESSSLFOTDLKLGSLRYQKLPSEDESG 1471
DB 361 NDASPLDPTTEDEKSDGSKLLPGKKSSESSSLFOTDLKLGSLRYQKLPSEDESG 420
QY 1472 TEESDNTPLDKDDKDKAEKGVKVERVPKSPHSAEPIRTFIKAKYLSLALLDKKSDSG 1531
DB 421 TEESDNTPLDKDDKDKAEKGVKVERVPKSPHSAEPIRTFIKAKYLSLALLDKKSDSG 480
QY 1532 VRSESSPNHSLHNEVADDSOLEKANLIELEDDSHSGRKGIPHSLSGLQDPIIARMSICS 1591
DB 481 VRSESSPNHSLHNEVADDSHLEKANLIELEDDSHSGRKGIPHSLSGLQDPIIARMSICS 540
QY 1592 EDKSPSCSLIASSPEENWACQKAYNLNRTPTVTTLNNSAPANRANQNFDEMEGIRE 1651
DB 541 EDKSPSCSLIASSPEENWACQKAYNLNRTPTVTTLNNSAPANRANQNFDEMEGIRE 600
QY 1652 TSQVILRPSSPNPTTIQENLKMTHKRSORSSYTRLSKDPPELHAAASESTGFGEER 1711
DB 601 TSQVILRPSSPNPTTIQENLKMTHKRSORSSYTRLSKDPPELHAAASESTGFGEER 660
QY 1712 ESIL 1715
DB 661 ESIL 664

RESULT 11
Q9H9E4 ID Q9H9E4 PRELIMINARY; PRT; 543 AA.
AC Q9H9E4;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ12811.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

SEQUENCE FROM N.A.
PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kinata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
RA Moriya S., Momiya M., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama K., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
```

```
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs."
RT Nat. Genet. 36:40-45 (2004).
DR EMBL; AK022873; BAB14285.1;
DR InterPro; IPR010993; SAM homology.
SQ SEQUENCE 543 AA; 59976 MW; FSD643D5A20C641D CRC64;

Query Match 31.4%; Score 2791; DB 2; Length 543;
Best Local Similarity 99.6%; Pred. No. 1.3e-128;
Matches 541; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1173 MLPOYCTTIKANKINGRVLAQCNDIDELKKNMNFQDWHLFRSTVLEWRNAESHVVPD 1232
DB 1 MLPOYCTTIKANKINGRVLAQCNDIDELKKNMNFQDWHLFRSTVLEWRNAESHVVPD 60
QY 1233 RFLSESSGPPHPEARRASHNELPHELSSQTPYTLNFSFEELNTGLDEGAPRHNL 1292
DB 61 RFLSESSGPPHPEARRASHNELPHELSSQTPYTLNFSFEELNTGLDEGAPRHNL 120
QY 1293 SWQQTTRTPSLSLNSQDSSIEISKLTDKQAEYRDAYREYIAQMSQLEGPGSTTISG 1352
DB 121 SWQQTTRTPSLSLNSQDSSIEISKLTDKQAEYRDAYREYIAQMSQLEGPGSTTISG 180
QY 1353 RSSPHSTYMGQSSSGSIHSNLEQKKGKQSEPKDGRKSFMLMKRGVDVYSSSGYST 1412
DB 181 RSSPHSTYMGQSSSGSIHSNLEQKKGKQSEPKDGRKSFMLMKRGVDVYSSSGYST 240
QY 1413 DASPLDPTTEDEKSDGSKLLPGKKSSESSSLFOTDLKLGSLRYQKLPSEDESGT 1472
DB 241 DASPLDPTTEDEKSDGSKLLPGKKSSESSSLFOTDLKLGSLRYQKLPSEDESGT 300
QY 1473 EESDNTPLDKDDKDKAEKGVKVERVPKSPHSAEPIRTFIKAKYLSLALLDKKSDSGV 1532
DB 301 EESDNTPLDKDDKDKAEKGVKVERVPKSPHSAEPIRTFIKAKYLSLALLDKKSDSGV 360
QY 1533 RSSESSPNHSLHNEVADDSOLEKANLIELEDDSHSGRKGIPHSLSGLQDPIIARMSICSE 1592
DB 361 RSSESSPNHSLHNEVADDSHLEKANLIELEDDSHSGRKGIPHSLSGLQDPIIARMSICSE 420
QY 1593 DKXSPSCSLIASSPEENWACQKAYNLNRTPTVTTLNNSAPANRANQNFDEMEGIRET 1652
DB 421 DKXSPSCSLIASSPEENWACQKAYNLNRTPTVTTLNNSAPANRANQNFDEMEGIRET 480
QY 1653 SQVILRPSSPNPTTIQENLKMTHKRSORSSYTRLSKDPPELHAAASESTGFGEER 1712
DB 481 SQVILRPSSPNPTTIQENLKMTHKRSORSSYTRLSKDPPELHAAASESTGFGEER 540
QY 1713 SIL 1715
DB 541 SIL 543

RESULT 12
Q7KVP5 ID Q7KVP5 PRELIMINARY; PRT; 1604 AA.
AC Q7KVP5;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DN CG30387-PC.
GN ORFNames=CG30387;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Spidroidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
```

RA Brandon R.C., Rogers Y.H., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA April J.F., Agbayani A., An H.J., Andrews-Pfankoch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Chertys J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Houten D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy C., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "the genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Calniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,
 RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
 RT melanogaster euchromatic genome sequence";
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RN [3]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Calniker S.E.,
 RT "The transposable elements of the Drosophila melanogaster euchromatin:
 RT a genomics perspective";
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RN [4]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celisner S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.,
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review";
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [5]
 RN SEQUENCE FROM N.A.
 RP FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]

RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AS003453; AAM70867.2; -
 DR HSSP; Q00420; 1AWC.
 DR InterPro; IPR002110; ANK.
 DR Pfam; PF00123; Ank; 10.
 DR PRINTS; PR01415; ANKYRIN.
 DR SMART; SM00248; ANK; 12.
 DR PROSITE; PS00088; ANK_REPEAT; 9.
 DR PROSITE; PS0297; ANK_REPEAT; 1.
 KW ANK repeat.
 SQ SEQUENCE 1604 AA; 176580 MW; 3259E2F4F183229A CRC64;

 Query Match 27.2%; Score 2417; DB 2; Length 1604;
 Best Local Similarity 35.8%; Pred. No. 1.4e-109;
 Matches 597; Conservative 286; Mismatches 552; Indels 232; Gaps 48;

 QY 1 MSYLISQSVINVEENIPALKALLE-KCKVDVERNECCGOTPLMIAAEQNLIEIVKELIK 59
 DB 1 MSLGHRALLQYIDNNDISGLRALDSRLHTIDDRDENATTVLMVAGRLTAFVREFLA 60
 QY 60 NGANCNLELDNWTALISASKEGHVHIVBEELKCGVNLHEDRMGGWTALMACYKQRTDV 119
 DB 61 RGADVQAEELDNWTALLCASRNGLDVLQVLLDHGAEEVHRDGGWTSMLWAAVYRGHTEL 120
 QY 120 VELLISHGANPSVTGLQSVYPIIWAAGRGHADIHLLLLONGAKNCSDKYGTTPLVWAA 179
 DB 121 VRLLLDKGADGNAHG-NYHLGALLWAAAGRYKDIIVELLVQRGAKVNVGKYGTTLVWAC 179
 QY 180 RKGHLCVCKHLLAMGADVDEQANSMTALIVAKGYTQSVKEILKRNPNVNLTDKQNT 239
 DB 180 RRGNEIVDTLLKAGANVTAGYSWTPLLVAAGGHTDCVSSILEKKPNVNLADKQMT 239
 QY 240 ALMIASKEGTEIVQDLDDAGTYVNIPIRSGDTVLIGAVRGHVEIVRALLQKYADIDIR 299
 DB 240 ALCIASREGFQDIAASLIAAGAYINIQDRGADTPLIHAVKAGHRTVVEALLKKGADVDIQ 299
 QY 300 QDNKTALYWAKEGNATVRILOQNPOTETCTKDGTEPLIKATKWRNIEVVELLDKG 359
 DB 300 GKDRKTAIYAVEKGHTPIVKLLATNPDESATKDGDTPLLRVNRNRLIEVHLLDRK 359
 QY 360 AKVSAVDKGDTPHIAIRGRSRKLAELLRNPKGRGLRPNKAGETPVNIDCSHQKSI 419
 DB 360 AKVTASDKGDTLHIAIRGRSRKLAELLRNPKGRGLRPNKAGETPVNIDSLHQYI 419
 QY 420 LTQIFGARHLSPTETDGMGLYDLYSSALADILSEPTMQPPICVGLYAWGSKSFLKK 479
 DB 420 LGQVFGARRLNTNDESGMLGYELYSALADVLSEPTLTATPTVGLYAKWGSKGSFLNK 479
 QY 480 LEDEMTFAGQIEPLFQFSLWLVFTL---LLCG---GLGLLFAFVHPNLGIAVLSLF 533
 DB 480 LRDEMNFARQWAEPPRTSGLLFVCLHVALLIGTVGLSTWSAV-----VGVSAAVGF 534
 QY 534 LALLYFFIVYFGRRREGESWNWAWLSTRLARHIGYLELLIKLMFVNPPPEQTTKA 593
 DB 535 LLLAYLLLAARYCNQ--MDMQWYSVQVHGLEKMTLRLLILOVAFCHPPG-POSDSQA 591
 QY 594 LPVRFLLFTDYNRLSSVGGTSLAEMTATISDACEREFGLFATRLFRVFKTE--DTQKKK 651
 DB 592 KPVRFHFAEANGASPT-GDGAHMLAALLDAIESHYGWLATRLYRAFRPKCLKVDVGWR 650
 QY 652 WKYTCCLPSFVFLFIIGCIISGITLAIFF-----RVPDKHLTVNAVLISASVGLAFV 706
 DB 651 WRMCCIPVLIFELAVLVTVTGISLTVAIFYTFADEKEKEHLV--ALYVIAAVMGTLIC 708
 QY 707 LNCRTWQVQLDLSLNSQRKLNAAASKHLKXSEGMFKVLCKEVELMARMAKTIDSTQN 766
 DB 709 THLVLAUVFVSLFTSHIRVLKRAV-----RSSESAPLTMLGAEVAVTDMVKCLDAFTNQ 764
 QY 767 QTRLVVVIDGLDACEODKVLQMLDVTVRVLFSS--KGFPIAIFASDPHIITKAINQNLSVL 824
 DB 765 QSRVGVVIDALDSCDTERILTILINAVQTLTSSPNRPVLLISVDPHVIKAAEANSRRLF 824

RL	Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).	
RN	[5]	
RP	SEQUENCE FROM N.A.	
RG	FlyBase;	
RL	Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.	
RN	[6]	
RP	SEQUENCE FROM N.A.	
RG	FlyBase;	
RL	Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AE003453; AAF46710.4; -	
DR	HSSP; Q00420; 1AUC.	
DR	InterPro; IPR002110; ANK.	
DR	Pfam; PF00023; Ank; 10.	
DR	PRINTS; PR01415; ANKYRIN.	
DR	SMART; SM00248; ANK; 12.	
DR	PROSITE; PSS0088; ANK REPEAT; 9.	
DR	PROSITE; PSS0297; ANK_REPEAT; 1.	
KW	ANK repeat.	
SQ	SEQUENCE 1626 AA; 179018 MW; 04525871BCDD393C CRC64;	
	Query Match 27.2%; Score 2417; DB 2; Length 1626;	
	Best Local Similarity 35.8%; Pred. No. 1.5e-109;	
	Matches 597; Conservative 286; Mismatches 552; Indels 232; Gaps 48;	
QY	1 MSVLISQSVINYVEENIPALKALLE-KCKQVDERNECGOTPLMIAAEQGNLEIVKELIK 59	
DB	23 MGSIGHRALLOYIDNDISGLRAILDSRHLTIIDRDENATVLMVAGRLTAFVREFLA 82	
QY	60 NGANCLELDLNNLTALISAKESGHHVHVEELLKCGVNLHRDGGWTALMWACYKGRDVT 119	
DB	83 RGADVQAEELDNNLTALLCASNHGHLVDVQLLDHGAEEVHRDGGWTSLMWAAVYRGHTEL 142	
QY	120 VELLISHGANPSVTGQYVYPIIWAAGRHADIVHLLONGAKVNCSDKYGTTPLVMAA 179	
DB	143 VRLLLDKGADGNAGH-NYHLLGALLAAGRGYKDIVELLVQRGAKVNVGDKYGTALVWAC 201	
QY	180 RKGHLECVKHLAMGADVDOEGANSMTALIVA KGGYTSQVKEILKRNPNVNLTDKDGNT 239	
DB	202 RRGVVEIVDLLKAGANVDTAGMYSTPLVLAAGGHTDCVSILEKKFNVNALDKGWT 261	
QY	240 ALMASKEGHEITEIVQDLLDAGTVYNIIPDRSGDVTLVIGAVRGHVEIVRALLQKYADIR 299	
DB	262 ALCTASREGQDIAASLIAAGAYINIQRDAGDTPLIHAVKAGHRTVVEALLKGHADVDIQ 321	
QY	300 GQNKALYAWKGNATVRDILQCNPDTEICTKGETPLIKATKRNLEIVVELLDKG 359	
DB	322 GKDKTAIYTAVEKGHTPIVKKLLATNPDLSEATKQDGTPLLRVNRNLEIVHLLLDK 381	
QY	360 AKYSADVKKGDTPLHIAIRGRSKLAELLRNPKDGLLYRPNKAGETPNIDCSHQSI 419	
DB	382 AKVTASDKRGDTCLHAMRARSKTIVEALLRNPKHSQLLYRANKAGETPNIDSLHQYTI 441	
QY	420 LTOIFGARHLSPTETDGMDGLYSSALADILSEPTMQPPICVGLYAWQSGSKFLKK 479	
DB	442 LGQVFGARRLNTNEDSEGMLGYELYSALADVLSEPTLTPTVGLYAKWGSKFLNKK 501	
QY	480 LEDEMTFAQOQIEPLFQFSLVFLVTL---LLCG---GLGLFAPTVHNLGIASLSF 533	
DB	502 LRDEMNFAQWAEPPRTSGLLFIVCLHVALLIGTIVGLTSWAV-----VGVSAAVGP 556	
QY	534 LALLYTFPIVYFGRRREGSWNAWVLSRLARHIGYLELLKLMFNPPELPEOTTKA 593	
DB	557 LLLAYLLLAAYRCNVQ---MDMQWAVSVQHGLEKRTLRLELILOVAFCHPPG-PQSDSQA 613	
QY	594 LPVRFELTDYRNLSSVGGETSLAEMTATLSDACEREFGLATRLFRVFKTE--DTQGGKK 651	
DB	614 KPVRFHFAEANSASPT-GDCAVAHMLAALLDAIESHYGWLATRLYRAFRPKCLKVDVGWR 672	
QY	652 WKYTCCLPSVFIFLFIIGCIISGITLLAIF-----RVDPKHLTVNAVLISIASVGLAFV 706	
DB	673 WRKMCCIPVILFELALVTVTGISTVAYFTFADEKEKEHILV--ALYVIAAVMGTLIC 730	
QY	707 LNCRTWQVLDLSLNSQRKHLHNAASKLHKLKSEGPMKVLKCEVELMARMAKTIIDSTQN 766	
DB	731 THLHVLAQVFSLFTSHIRVLKRAV---RSSESAPLTMLGAEVAVMTDMVKCLDAFTNQ 786	
QY	767 QTRLVIIIDGLDACEQDKVQLMDLTVRVLFS--KGFPIAIFASDPDHIIIIKAINONLSVL 824	
DB	787 QSRLVGVIDALDSCDTERILTLLNAVQTLTLLSPNREFVILLISVDPHVIAKAAEANSRRLF 846	
QY	825 RDSNTNGHDYMRNIVHLPVFLNSRGLSNARKFLVTS-----ATNGDVPVPCSD--TTGIQED 877	
DB	847 TEGGIGGHDPLNMLVHLPVYLNQSLGRKVKVQRAQMTALLFKRSGGDIYQDGDGFLGHSVS 906	
QY	878 ADDRVSQNS--LGEMTKL-----GSKTALNR-----RDYRRRQVQRTITRQMSF 920	
DB	907 A-RRLSNASEIISQEKLRGARGGKKLRLSSESVASSTGSLNHLRGQNPQTV-----L 960	
QY	921 DLTKLLVTEWDFSDISPTQMRLLNIVSVTGLLRANQISFNWDRLASINLTQWPYRT 980	
DB	961 DLSRIVLTDYDFSDVNPFRMRLLMNVITYITVRLKAFQIEFYSWYRLSSWINLTQWPLRA 1020	
QY	981 SWLILVLEE--TEGIPDOMTLKTIYERISKNTPTTKDVEPLLEIDGDIRNFVFLSSRTP 1038	
DB	1021 SMIVLHDDFMDSNADESLSQSVYEKLRPKLAYLREAAPLLELDRDKLDAFLQHLKS 1080	
QY	1039 VLVAEDVKVFLPCTVNLDPKLRRIIAD-----VRAAREQISIGGLAYPPLLHE 1087	
DB	1081 DLLVADLRIFLPTINLDVLRKVLKEDQQTIEDEGSLVIQARPSVSNTRMQPF----- 1134	
QY	1088 GPPRPSGYSQPPSCVSCSTSFNGPPAGGVVSPQPHSY-----YSGMTG 1131	
DB	1135 ----APTYY-----VPSQAYPPYQFQNEYFANELSRNLSTSTE 1171	
QY	1132 PQHPFVNRGSGPARGPVV--LLNSLNVDAVCEKLEIGLDOSMLPQYCTTIKKANINGR 1189	
DB	1172 PVTPLINSDFSGDDILOTKLTDLTVEGVISLLDRIEDM-KPALFKLAPVLAENAINGR 1230	
QY	1190 VLAQCNIDELKEMNMNFGDWHLFRSTVLEMRNAESHVVP-----EDPRFISESSSGP 1242	
DB	1231 VLKCHDMPDKSVLGLSFGHWELFRLLITTLRECEK--LPRKQORQQQPGALEAPSVP 1288	
QY	1243 APHG-----EPARR--ASHNELPHE-----LSSQ---TPYTLNFSPELNTLGLDEGA 1286	
DB	1289 MIKDVTDALMQPPRESLSRKNVSHMEKQVTLFEEQMCITQTLNEEAYE-DVASSERPS 1347	
QY	1287 PRHSNLSWOSQRTPTSLSSL-----NSQSSIEISKLTDKVQ---AEY-RDAYREYI 1335	
DB	1348 PTGEMLAAVAQQLAPIRESSSEFSGSDQKQGVKISNNNNNNVYLHAEYNSVSSHSL 1407	
QY	1336 AQMSQLEGPGSTTISGRSSPH-----STY-----YMQSSSSG-----SI 1371	
DB	1408 QSLSTLVGAPVGHGSGGSHLHGNGDLSDSLMLHVDVSVFGGGGGYHRAQRQISI 1467	
QY	1372 HSNLEQEKGD-----SEPKDDGRKSFMLKRGDV-----IDYSS 1406	
DB	1468 SSELLHESKLDTPSAGAAMTTPLLGASGVPAAP-SGRESILKQQGSVKADKRVISIQMAT 1526	
QY	1407 SGVSTNDASPLDP-----ITEDEKSDQSGSKLL-----PGKK-----SSERSSLFQTDL 1451	
DB	1527 SNNNNNSKLTIPNVEYVSERQEVQAGKRLTKPPGPRPASLIITRDNDSQQL-- 1584	
QY	1452 KLKSGRLRYCKLPDSDESGTEESDNTPLLKDDKDKRKAEGKVERVPK 1498	
DB	1585 -LRSSVDYDDVEAQEHRT---TIRTLLEQEEBESAPFVFTVRK 1626	
RESULT 14		
Q9W210	PRELIMINARY; PRT; 1678 AA.	
ID	Q9W210	
AC	Q9W210;	
DT	01-MAY-2000 (T-EMBLrel. 13, Created)	
DT	01-JUN-2003 (T-EMBLrel. 24, Last sequence update)	
DT	01-MAR-2004 (T-EMBLrel. 26, Last annotation update)	
DE	CG30387-PB.	
GN	ORFNames=CG30387;	

OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Fandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Baldwin D.,
 RA April J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
 RA Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Follard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirskaas R., Taber P., Pfeiffer B.D., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: Release 3 of the *Drosophila*
 RT *melanogaster* euchromatic genome sequence."
 RL Genome Biol. 3:RESEARCH0073-RESEARCH0079(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskaas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.;
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
 RT a genomics perspective."
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whittied E.J., Bayraktaroglu L., Berman B.P.,

RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review."
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [5]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003453; AAF46711.4; -.
 DR HSPF; P42773; I1HB.
 DR InAct; Q9W210; -.
 DR FlyBase; FBgn0050387; CG30387.
 DR InterPro; IPR002110; ANK.
 DR Pfam; PF00023; Ank; 10.
 DR PRINTS; PR01415; ANKYRIN.
 DR SMART; SM00248; ANK; 12.
 DR PROSITE; PS50088; ANK_REPEAT; 9.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 KW ANK repeat.
 SQ SEQUENCE 1678 AA; 184516 MW; D75A23F12B545DFC CRC64;
 Query Match 27.2%; Score 2417; DB 2; Length 1678;
 Best Local Similarity 35.8%; Pred. No. 1.5e-109;
 Matches 597; Conservative 286; Mismatches 552; Indels 232; Gaps 48;
 QY 1 MSVLSQSVINYVEENIPALKALLE-KCKDVSERNECGQPLMTAAEQGNLEIVKELIK 59
 DB 75 MGSUHRALLQYIDNNDDSLGRLSDSRHLTDIDRDENATTVLVVAGRGITAFVREFLA 134
 QY 60 NGANCNLEDLNTALISASKEGHVHIVVEELLKCGVNLHEDMGWGTALMWACYGRDIDV 119
 DB 135 RGADVQAEEDLNTALLCASRNGHLDVVQLLDHGAVEHEDMGWTSLSMAWVYRGHTL 194
 QY 120 VELLSHGANSVTGLQSVVPIIWAAGRHADIVHLLONGAKVNCSDKYGTTPVLWAA 179
 DB 195 VRLLLDGADGNAHG-NYHLGALLWAAAGRYKIDIVELLVQRGAKVNVGDKYGTTLVWAC 253
 QY 180 RKGHLECKVHLLANGADVQEGANSMTALIVAVKGYTQSVKEILKRNPNVNLTKDQNT 239
 DB 254 RRGVNEIVDTLLKAGANVDTAGMYSWTPLLVAAAGGHTDCVSSILEKKNVNLADKGMT 313
 QY 240 ALMTASKEGHTPIVQDLIDAGTYVNIPIRSQDVTVLIGAVRGHVEIVRALLQKVADIDIR 299
 DB 314 ALCTASREGFDIAASLIAGAYINIQRGADTPIHAKAGHRTVVEALLKGHADVDIQ 373
 QY 300 QDNKTALYWAVERKGNATMVRDILOCPDTEICTKGSTPLIKATKMNIEVVELLLDKG 359
 DB 374 GKDRKTAIYAVEKGHPTIVKLLLATNPDLSESATKDGDTPLLRVNRNLEIVHLLDRK 433
 QY 360 AKVSADVCKGDTPLHAIARGSRKLAELLNPKDGLLYRPNKAGETPNYDCSHKSI 419
 DB 434 AKVTASDKRGDTCLHIAWRAKSKTIVKALLRKHSLLYRANKAGETPNYDLSHQTKI 493
 QY 420 LTQIFGARHLSPTEDTGDMGLGYDLYSSALADILSEPTMOPPICVGLVAQMGSGKSLKK 479
 DB 494 LGQVFGARRLNTNEDSEGMLGYELYSALADVLSEPTLTPITVGLYAKMGSGKSLNK 553
 QY 480 LEDEMKTFAGQIIEPLQFQSWLIVFLTL---LLCG---GLGLLFAFTVHPNGLIAVSUF 533
 DB 554 LRDEMNFRQWASPPRTSGLLFIVCLHVALLIGTVGLTSVAV-----VGVSAAVGF 608
 QY 534 LALLYIFPIYVFGRRREGESWNWVLSFLARHIGVLELLLMFVNPPELPEQTTKA 593
 DB 609 LLLAYLLAAVRYCNYQ--MDMOWAYSQVHQHKEKMTLRLILQVAFCHPFG--PQSDQSA 665
 QY 594 LPVRFLFTDYNRLSSVGGETSLAEWIAATLSACERFGLATRLFRVFKTE--DTQGGKK 651

```
Db 666 KPVRFHFAEANSAPT-GDCAVHMLAALDALSHYGLWATLYRAFRPKCLKVDVGWR 724
QY 652 WKTCCLPSPVIFLFIGIISGITLAI- - - - -RVDPKHLTVNAVLSIASVWGLAFV 706
Db 725 WRMCCIPVILFELALVTVTGISTVAVFTFADEKEKEHILV-ALYVIAAVMGTLIC 782
QY 707 LNCRTWVVLDSLNSQKELHNAASKLHKLSESGPMVKLKCVELMARMAKTIDSFQTN 766
Db 783 THLVLAKVEVSUFTSHIRVLKAV- - - - -RSSSAPLTMIGAFAVWMTDMVKCLDAFTNQ 838
QY 767 QTRLVIIIDGLDACEQKVQLMDLTVRVLFS- - - - -KGPFIALFASDPHIILKAINQNSVL 824
Db 839 QSLVGVIDALDSCDTERIILTNVQTLSSPNRPVLLISVDPHVIKAAANRRRLF 898
QY 825 RDSNINGHYMNVHLVPLFNSRGLSNARKFLVTS- - - - -ATNGDVPCSD- - - - -TTGIQED 877
Db 899 TEGGIGCHDFLRNLVHLVPLVQLNSGLRKVQRAQMTALLPKRSGGGDYQTDGPTLGHVS 958
QY 878 ADRVQSNS- - - - -LCEMTKL- - - - -GSKTALNR- - - - -RTYRRQMQORTITQMSF 920
Db 959 A-RRLSNASEIISQEKLRGAPARGGGKRLRLSESVASSTGSLNHLRGLQNPQTV- - - - -L 1012
QY 921 DLTKLAVTEDFDIDSPQWRRLLINIVSVTGRLLRANQISFNWDRLASWINLTEQWPYRT 980
Db 1013 DLSRIVLTDYFSDVNPNSRNLNVIYIIVRLKAFQIEFYSWRLSSWINLTEQWPLRA 1072
QY 981 SMILYLEE- - - - -TEGIPQWMTKTYIRISKNIPPTKDVPELLEIDGIDRNFVFLSSRTP 1038
Db 1073 SMIVLHHDQPMDSNADESLSQSVYEKRLPKLAYREAAPLELDRDKLDAFLQLHKS 1132
QY 1039 VLVARDVKVLPCTVMDPKLRIIAD- - - - -VPAAREQISIGGLAYPPLPLHE 1087
Db 1133 DLLVADLRIEPLPILNDPLRKLVEDQDOTIEDEGLSVIQAAPSVSNTRQPP- - - - - 1186
QY 1088 GPPRAPSGYSQPPSVCSSTFNGFPAGGVVSPPHSSY- - - - -YSGMTG 1131
Db 1187 - - - - -AFTY- - - - -VPSQAPYPPYQMFONEYPANELSRNLSTSTE 1223
QY 1132 PQHPFYNRGSGPAPGV- - - - -LLNSLVNAVCEKLOIEGLDQSMPLPOYCTTIKANINGR 1189
Db 1224 PVTPLNSPDSFGDDIQLKTLDTAVGVISLLDIEDM-KPALPKLAPVLRENAINGR 1282
QY 1190 VLAQCNIDELKEMWNFQDWHLFRSTVLEMRNAESHVP- - - - -EDPRFLSESSGP 1242
Db 1283 VLKHCMDPLKSLVGLSGFHWELFRLLITTLRECER- - - - -LPRKOORQQQPGALEAPSNVP 1340
QY 1243 APHG- - - - -EPARR- - - - -ASHNELPHE- - - - -LSSQ- - - - -TPYTLNFSFEELNTLGLDEGA 1286
Db 1341 MIKQVTDALMQPPRESLSRKNSVSHMEKQVTLSEQMLCGTLQTLNEAYE-DVASSERPS 1399
QY 1287 PRHSNLSWQSQRTRTPSLSL- - - - -NSQDSSIEISKLDKVQ- - - - -AFY-RDAYREYI 1335
Db 1400 PTGEMLAAVAQLQAPITRESSEFSGSDQKQYGVKISNNNNNQYLHAEYNSVSSHSL 1459
QY 1336 AQMSOLEGGPGSTISGRSPH- - - - -STY- - - - -YMGQSSSG- - - - -SI 1371
Db 1460 QSLSTLVGAPVHGGSGGSHLHNGNDLSDSLDLMHVDVSFVGCGGGGYHRAQRQISI 1519
QY 1372 HSNLEQKRGKD- - - - -SEPKDDGRKSPFLMKRGDVP- - - - -IDYS 1406
Db 1520 SSELLHESKLDTPSAGAMTTPLLGASGVFAP-SGRESILKQOGSVKADKRVSIQOMAT 1578
QY 1407 SGVSTNDASLDP- - - - -ITEDEKSPQSGSKLL- - - - -PGKK- - - - -SSESSSLFQIDL 1451
Db 1579 SNNNNNSKLTNPVENVVSERQEVQAGKRLTTKPPGPPRPASLIITRNDNSQFOL- - - - - 1636
QY 1452 KLKSGSLRYOKLRSDEDESGETESDNTPLDKDKRKAEGKVERVPK 1498
Db 1637 -LRSSSYDDVDEAQEHRT- - - - -TIRTTLEQQEESAPFVFTVRK 1678
```

RESULT 15

```
Q7PZY2
ID 07PZY2 PRELIMINARY; PRT; 1354 AA.
AC Q7PZY2;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE EbiP9166 (Fragment).
GN Name=ebiG9166; ORNames=ENSANGG00000006867;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data
CC EMBL; AA01008986; EAA00198.1; -.
DR HSSP; O75832; IUOH.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; Ank; 11.
DR PRINTS; PR01415; ANKYRIN.
DR PROSITE; PS00088; ANK_REPEAT; 9.
DR PROSITE; PS0297; ANK_REPEAT_REGION; 1.
DR ANK repeat.
FT NON TER 1354 1354
FT NON TER 1354 1354
SQ SEQUENCE 1354 AA; 148250 MW; E0814F869C911280 CRC64;

Query Match 26.2%; Score 2328.5; DB 2; Length 1354;
Best Local Similarity 39.0%; Pred. No. 2.4e-105;
Matches 525; Conservative 243; Mismatches 440; Indels 137; Gaps 31;

QY 1 MSVLISQSVNVYVEENIPALKALL-EKCDVDNERCEGOTPLMIAEQNLBIVKELIK 59
Db 3 MSLGHRSLQYLETDDTLGSLKSLFGLRHLQVDDRDENNITVLVWASGRGATHFVELLA 62
QY 60 NGANCNLELDLNNWTALISAKEGHIVIEELLKCGVNLHRDMGWTALMWACYKGRD 119
Db 63 RGADVQAQDLDSWTALHFAAKAGHVIGVELLLDNGAELEHRDMGWTALMWGSKGHTSV 122
QY 120 VELLSSHGANPSVTGLOYSYVPIIAAGRGHADIIVHLLIQ-NGAKVNCSDKYGTPLVWA 178
Db 123 VALLLQRGADVQAHG-NYHLNPLLWASGRGHTIEVELLVNTGKGVNVDGKGTPLVWA 181
QY 179 ARKGHLECVKHLAMGADVDOEGANSMTALIVAVKGGYTQSVKEILKRNPNVNLTKDGN 238
Db 182 CRKGSABIVDVLKAGANVDTAGWYSWTPLLVAVSGGFQECVSLLLERKPNVNLKXGM 241
QY 239 TALMTIASKEGHTIEVDLLDAGTYVNIPIRSGDTVLIGVRGGHVIEIRALLQKYADIDI 298
Db 242 TALSIAACREGLTEIASALIAAGAYLVNQDRAGDTPLINAVKGGHRSVVEILMKRHVDVI 301
QY 299 RGDNNTALYAVKGNATWDRDILQCNPDTEICTGDTPLIKATKMRNIEVVELLDK 358
Db 302 QGKDKKTALYTAVEKGHTAIVKILQSNPDLSELTGDTALLRAVRNRLVQMLER 361
QY 359 GAKVSAVDKGGDTPHIAIRGRSRKLAELLRNPKDRLLYRPNKAGETPNYDCSHOKS 418
Db 362 KAKVGATDKRGDTCLHVAWRARSKAIVEALLSNPKYQLLYRSNKEGETPYALDTHQKT 421
QY 419 ILTQIFGARHLSPTETDGMGLGYDLYSSALADILSEPTMQPPTCVGLYAGWGSKGFLLK 478
Db 422 ILQVFGNRRNLANEDSEGMLGYLYSSALADVLSEPTLTPTITVGLYAKWGSKGFLLT 481
QY 479 KLEDEMTFAGQOIEPLFOFSMLIVPLTLLCGGLGLLFAFTVPHNL-GIANSLSFALL 537
Db 482 ELRDEKNFNAHSWSEPPIDASWMLFFLISHLVLVIGTVVGLATWVWGVMTGATLLVLI 541
QY 538 YIFFIVIFYGRRREGSGSNWAVLSLRLARHIGYLELLLLKLMFVNPPPEQTTKALPVR 597
```

Db 542 YFTDILLKPLRR--YDLEWYSLNYGLSRKLGRLILQVAFCHPPG-POSQOQMEVR 598
Qy 598 FLFTDYNRLSVGGSTSLAEMIATLSDACERFEGFLATRLFRVFKTEDTQ--GKKWKKT 655
Db 599 FHFAEASG-AAPNGDAVALMLASLFDALIAHYGSLATGLYRAFPPKPKATGGWKWKOM 657
Qy 656 CCLPSFVIFLFIIGCIISGITLLAIFRVPDKHLTVNA--VLISIASVVGGLAF---VLNC 709
Db 658 CCMP--VVLMPFELGML--GILATASLSIVYSEYGLEGREETAVAIYVLLGILLAGAIANL 713
Qy 710 RTWQVLDLSLNSQKRLHNAASKLHLKSGFPMKVLKCEVELMARMMAKTIDSFTQOTR 769
Db 714 HAWSKLIGALFMSQCKHLKRAFNS-----NEAAPTALGAEVSLMTDMVRCCLDAFTGQOSR 769
Qy 770 LWIIGLDACQDQKVLQMLDTVRVLF--KGPFAIFASDPHIIIIKAINONLNSVLKDS 827
Db 770 LVGVVDALDSCDTERLTILNAVQTLTLLSGPQRPFFVLLAVDPHVIKAAEANSRLFTTEG 829
Qy 828 NINGHDYMRNIVHLVPFLNSRGLSNARKFLVT--SATNGDVPCSDTTGTGIOEDAD----- 879
Db 830 GIGGHDFLRLNVLPLVYQLNSGLKRVQRAQNTAMSAVRRAMP-----DLSRDDQPHLGH 884
Qy 880 ----RRYQNS--LGEWTKLGSKTALNRDTRYRRQM-----RTITROMSFDLT 923
Db 885 SASVRLRLSNASEIMSSQEKRAMPSRSSRGGSKLRVSDSIASSIGSNLHKLQNPVVDLS 944
Qy 924 KLLVTEQWFDISPTQWRLRLNIVSVTCRLLRANQISFNWDLASWINLTEOWPVRTSWL 983
Db 945 KIIITDDYFSDVNPFRMRRLMNVYIIVRLKAFQIDFSWYRLSSWINLTEOWPLRSMI 1004
Qy 984 ILYLEETEG---IPDQWTKTYERISKNIPTTKDVEPLLEIDGDIRNFVFLSSRTFVL 1040
Db 1005 V--LEHDQGBDSFDDSVSLQAVYDKVRPKLTCLREAAALLDLDRDKLDAFLQLHKSXL 1062
Qy 1041 VARDKVFLPCTVMDPKLREIADVRAREQ-----ISIGGLAYPPLPLHEGPPRAPSGY 1096
Db 1063 LVSLRIFLPPTIINLDPYLRKVLKEDQQALEDGEGVYGTNALMGEPLP----- 1110
Qy 1097 SQPPSVCSSTSFNGPFAGGVVSPOP---HSSYYSG-----MTGPQHPFYNRSGGPAP 1145
Db 1111 -KPGSV-----GGPIITEHPHEQHTLGSNSGRSKTTTGALQKSSNGLATSP 1160
Qy 1146 GPV-----VLLNSLVDAVCEKIQIEGLDQSMPLQYCTTIKKANINGRVLAQCNIDEL 1199
Db 1161 -PIDIDLSNVQLSSLTVEQLIELLEKVNDLKPAM-ERTAPILRENAISGRVLMYCNLEEL 1218
Qy 1200 KKENMHFGDWHLFRSTVLENRNA-----ESHVVEDP 1232
Db 1219 KSVLRLSFGHWEMPKLLVLTALREASVTQPASRKLSTKTTSPAKGTNDSVENQEPFAQSTP 1278
Qy 1233 RFLSESSSGPAPHGEPARRASHNEL 1257
Db 1279 PFGSSTSS-----FQPIRQKSONLL 1298

Search completed: February 5, 2005, 19:34:11
Job time : 202.5 secs

This Page Blank (uspto)